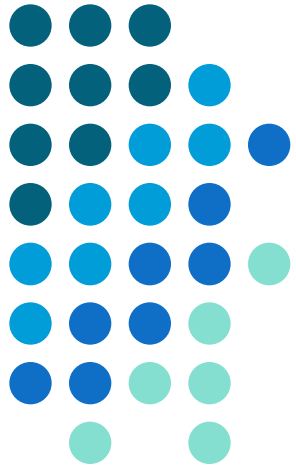
Data Validation, Verification and Usability

Quality Management Training Module

Marion Kelly, US EPA Office of Water
Dale Rushneck, Interface, Inc.
William Telliard, US.EPA (Retired)
Lynn Walters, CSC
December 7, 2010



Acknowledgements

We thank the following EPA offices and individuals for sponsoring and organizing this training:

- Great Lakes National Program Office
  - Louis Blume, QA Manager
- Region 5
  - Kevin Bolger, QA Manager
  - Amberina Khan, Regional QA Chemist
- GLRI QA Technical Conference Steering Committee
  - EPA: Louis Blume (GLNPO), Kevin Bolger (Region 5), Mike Johnson (OSWER), Marion Kelly (OW), Amberina Kahn (Region 5), Deborah Lamberty (GLNPO), Carol Lynes (Region 2), Pranas Pranckevicius (GLNPO), Donna Ringel (Region 2)
  - CSC: Molly Middlebrook Amos, Elizabeth Benjamin, Judy Schofield, Lynn Walters

EPA United States Environmental Protection Agency

Logistics - Handouts

- Location to retrieve handouts
  - *http://epa.gov/greatlakes/qmp/qmtraining.html*
- Course-specific handouts
  - Glossary of data review and related terms (HO #1)
  - Example Guidelines and Checklists
    - Sample Control Center Data Review Checklist (HO #2)
    - Data Review Guidelines for PCBs By Method 1668A (HO #3)
    - OW CCH Effluent Guidelines Data Review Checklist for Dioxin/Furans via Method 1613B (HO #4)
    - Method 625 Data Review Checklist for the 2009 EPA EAD Study of Detection and Quantitation Limit Procedures (HO #5)
- Handouts for all courses today
  - Helpful Hints for Course Participants, List of QA-related References, Glossary of QA-related terms, QA Cheat Sheet, What Does our Quality System Cover?

Logistics - Questions

- Web participants may submit questions at any time
- Breaks for questions are scheduled throughout the course

Training Goals

Teach participants:

- The importance of addressing data verification and validation needs and strategies during early stages of project planning
- Terminology associated with data review
- Basic data verification and validation principles and important questions to ask
- What's necessary to apply data to an environmental decision

Agenda (What's Covered)

- Part 1: Introductory Concepts/ Building Blocks
- Part 2: Data Verification and Validation
- Part 3: Data Quality Assessment
- Part 4: Information Correction
- Part 5: Conclusions & Closing Remarks

Plan → Evaluate & Document → Correct

What's NOT Covered

- Step-by-step instructions for performing a data review
  - Highly method-specific, performed by experienced chemist or biologist
- Detailed instruction on systematic planning or preparing a QA Project Plan
  - Separate training is available on planning & documentation
  - Focus here is on important data review aspects to consider when planning your project, documenting your plans in the QAPP, implementing your data review plans, and documenting results of your review

Agenda- Part 1

Part 1: Introductory Concepts & Building Blocks

- Terminology
- Why data review, verification, and validation are necessary
  - Data Quality Act
  - EPA's Quality System & QA Project Plans
- Existing Data
- Building it into the project during the planning stage
  - Systematic Planning and Controlling Error
  - Data Quality Indicators & QC Measures
  - Sampling Design Considerations
  - Analysis Considerations
  - Lab Considerations
  - Reporting Consideration (elements and format)
- Examples

Estimated Time for Part 1: 50 minutes *(followed by a short break)*

Data Review, Verification, Validation & Data Quality Assessment Terminology

- Meaning of terms vary by organization
  - Generally differ in rigor & independence of the check
  - Important to explain what you mean and how you will do it
- Data Review
  - Typically means “to examine and inspect” (Webster)
  - Many organizations refer to this as encompassing data verification and validation
  - Others use this to mean only internal/in-house data verification before data are submitted or released
  - *In this training:* Includes data verification and validation

Terminology

Processes covered include:

- Internal verification
- Independent verification
- Verification of completeness
- Verification of compliance with method, procedural, or contractual requirements
- Determining the analytical quality of a specific data set
- Determining if the data meets the project's quality objectives and can be considered sufficient for its intended use

EPA OEI Terminology

- **Verification:** Evaluating the completeness, correctness, and conformance/compliance of a specific data set against method, procedural, or contractual requirements
- **Validation:** An analyte- and sample-specific process that extends the evaluation of data beyond method, procedural, or contractual compliance (i.e., data verification) to determine the quality of a specific data set
- **Data Quality Assessment:** A statistical and scientific evaluation of the data set to determine the validity and performance of the data collection design and statistical test, and to determine the adequacy of the data set for its intended use

Source: *EPA Guidance on Data Verification and Data Validation* (http://www.epa.gov/quality/qs-docs/g8-final.pdf)

Simpler Terms

- Are the data present; i.e., is there a result (or non-detect) for every analyte in every sample and a result for every analyte in every QC sample?
- Do the sample numbers match up with the Traffic Report or Chain of Custody
- Were all tests performed in the proper order?
- Is everything consistent?

Data Verification

- Do the data make sense?
- Is there anything weird?
  - e.g., amenable cyanide results are higher than total cyanide results

Data Validation

Data Review

- Do the data support the decision to be made?
- Are the data consistent with other data?

Data Quality Assessment

EPA United States Environmental Protection Agency

Terminology Summary

- Don’t assume you know what someone means when they say they will review, verify, validate, or assess data quality
  - Although definitions exist, they are not universally used by all organizations or even individuals within an organization
  - Ask them to explain and document what they mean and how they will do it!

Why is Data Review Necessary?

- Data Quality Act (2001)
  - aka “DQA”, but don’t confuse with Data Quality Assessment!
  - aka Information Quality Act
- EPA Quality System
- Stakeholder Perception and Confidence
  - Congress
  - Public
  - Partner organizations (including Canada, for the Great Lakes)
  - Regulated entities
  - Regulatory/Control Authorities

Data Quality Act

- Requires government-wide standards for “ensuring and maximizing the **quality, objectivity, utility, and integrity** of information (including statistical information) disseminated by Federal agencies”
  - OMB has oversight responsibility
  - OMB *Information Quality Guidelines* define “quality” as encompassing objectivity, utility, and integrity
  - OMB Guidelines require:
    - **Pre-dissemination review** of information by Agencies
    - Means for public to submit **Requests for Correction** (RFC)
- References
  - Act: Section 515(a) of the U.S. Treasury and General Gov’t Appropriations Act for FY 2001 (P.L.106–554; H.R. 5658)
  - OMB IQGs: 67 FR8452; February 22, 2002

Objectivity - defined by 2 measures

1. **Presentation**: Is the information being presented
   - Within a proper context?
   - In a way that is accurate, clear, complete, unbiased, and transparent?
2. **Substance**:
   - Is the information accurate, reliable, and unbiased?
   - Were the data and results developed using sounds statistical and research methods?
   - Are “influential” scientific results capable of being reproduced?

Utility

- Refers to the usefulness of the information to intended users
- Agencies must consider uses from the perspective of the Agency and the public
- Agencies must address reproducibility and transparency relevant to the usefulness of information from the public's perspective

Integrity

Refers to the **security** of information

- Protects the information from unauthorized access or revision
- Ensures the information is not compromised through corruption or falsification

EPA Implementation of DQA

- EPA issued ***Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility and Integrity of Information Disseminated by the Environmental Protection Agency***

  EPA/260R-02-008; October 2002

  epa.gov/quality/informationguidelines/documents/EPA_InfoQualityGuidelines.pdf
- Articulates EPA's position on use of quality data in making decisions
- Underscores EPA's commitment to disseminating information that is accurate and useful to the intended user

Data Quality Act Summary

- Nearly all information EPA and its partner organizations generate or use is subject to the Data Quality Act
- DQA compliance requires all data/information to be:
  - Transparent
  - Accurate
  - Reliable
  - Unbiased
  - Useful
  - Secure
- Cannot comply unless you review your data!

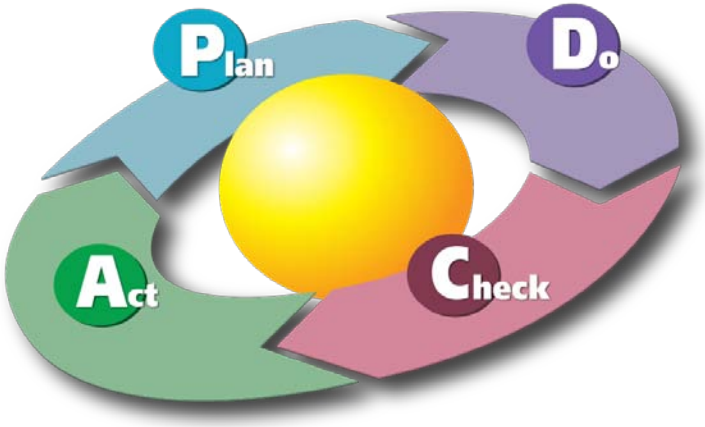
Cradle-to-Grave Data Quality Management

- **Planning** – what should be specified prior to data generation?
- **Data Reporting** – What level and types of data do data gatherers need to report?
- **Data Review** – What types of assessments should be performed before using the data?
  - **Includes using existing data** – How do I assess the quality of existing data?

Good News: EPA's Quality System addresses all these

EPA Quality System

- Complements the Information Quality Guidelines
- Primary goal: Ensure that environmental data are of sufficient quality and quantity to support its intended use

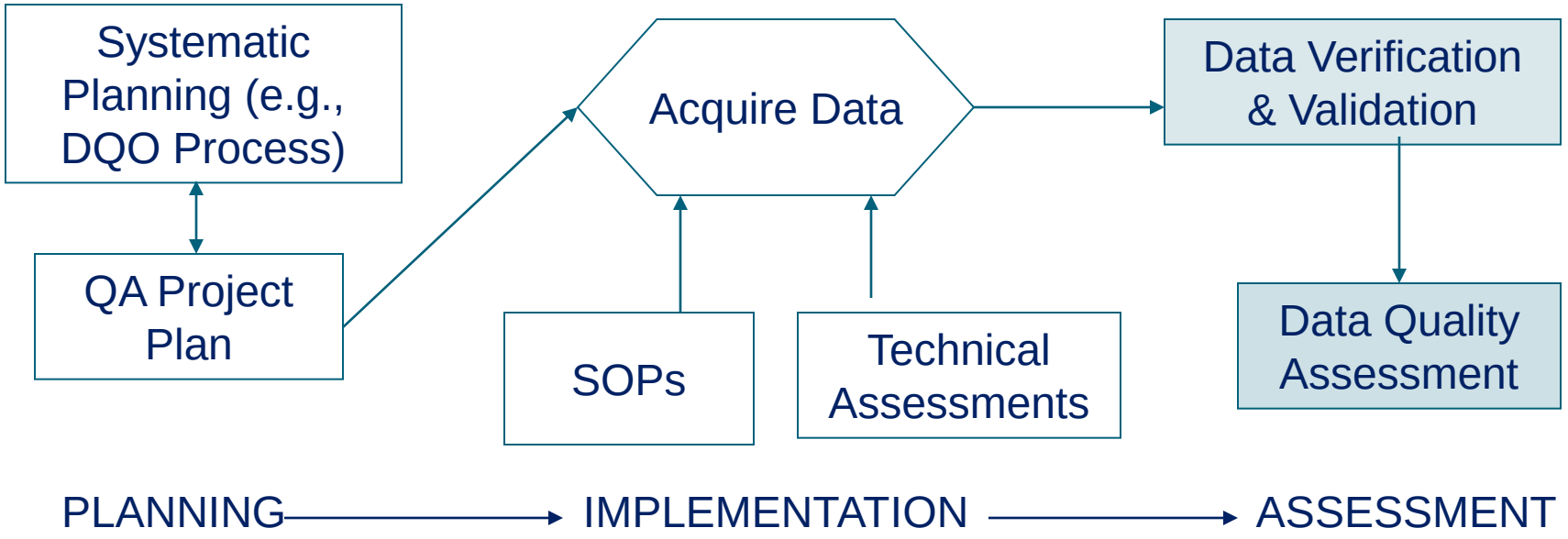


Quality System (continued)

- Covers data collection, evaluation, and use
- Decentralized so each organization within EPA designs, implements, and manages its own quality system
- Divided into 3 major components
  - Policy
  - Organization/Program
  - Project
    - This training focuses on the “project” component

Project Component of EPA's Quality System

- Data Review is a critical component of EPA's Quality System at the project level.
- Begins with systematic planning to address your needs



Source: *Overview of the EPA Quality System for Environmental Data and Technology*, Figure 8

EPA United States Environmental Protection Agency

Document Planning Decisions in a QA Project Plan (QAPP)

- Mandatory planning document for any data activity
- Describes how project data and information will be collected, analyzed and assessed.
  - A **'blueprint' for how each project will be implemented**
  - Must address all aspects of the project including planning, sampling, analyses, quality control; and data review**, validation, and** data quality assessment
- Must be in place before data collection or use begins
  - Contractors, IA partners grantees, and EPA staff cannot start work without an approved QAPP in place

References:

EPA QA/R-5, EPA Requirements for QA Project Plans
EPA QA/G-5 series, Guidance for QA Project Plans

What goes into the QAPP?

The results of your systematic planning process:

- The project objectives
- The quantity and type of data needed for the project and for the decision to be made, and how this need was determined
- How the data will be used to support the project objectives and decision
- The criteria for determining the quality of the data and how those criteria were developed
- How, when, and from where data will be obtained, including existing data from secondary data sources
- How the data for the project will be analyzed, evaluated, and assessed (e.g., field or laboratory quality control operations, audits, technical assessments, models)

QAPP – Pertinent Elements (of the 24)

- **Problem definition**: decision to be made/hypothesis to be tested
- **Project description**: what's to be done
- **Quality Objectives and Criteria**: what are the data quality objectives (DQOs) for the project
- **Sampling design**: must support the decision to be made and DQOs
- **Analytical method(s):** What procedures will be used and what performance standards will you use, and what will be done when there is a discrepancy
- **Quality Control**: what QC strategies and control limits will you use and how often
- **Data review and validation**: How will you determine if data are valid
- **Data usability assessment**: How you will determine if data support decision?
- **Reporting**: How will you document all of the above and who will you report it to

Primary vs. Secondary Data

- Primary Data
  - Collected by EPA (or under its direction) for a specific purpose associated with the decision at hand
- Secondary (existing) Data
  - Use of existing data that were not directly generated by/for EPA to support the decision at hand
  - Most often overlooked when planning data verification/validation/quality assessment strategies
- All data, regardless of source, must be of known and documented quality

Existing (Secondary) Data

Can be:

- The source of all your project's data or information
- Part of the data or information used in your project
- Used only to plan your project
  - Identify data gaps and new data needs
  - Estimate error and design your sampling and analysis strategy accordingly
- Regardless of how you use it, and how much you use, you need to evaluate the quality of it relative to your use
  - *We'll discuss this more in Part 2 of the presentation*

Use of Existing Data

- Advantages
  - Cost effective
  - Quick & easy
  - Sampling and analysis not required
    - Solves sampling access problem
    - No lab hassles
  - Possible broader range of information over time and space
- Disadvantages
  - May not be consistent with data needs
  - May be incomplete
  - May be difficult to assess quality

Primary and Secondary Data Examples

Primary Data

- Any data collected by EPA or under EPA's control (e.g., contract, grant, etc)
- Measurement results on environmental samples (e.g., field or lab data re physical, chemical, or biological characteristics)
- Data on the physical location of such samples (e.g., latitude, city, depth)
- Field or lab data used to assess performance of treatment systems or technologies
- Financial info supporting development of rules, regs, or guidance
- Engineering Process Data
- Models and results produced from models developed by/for OW for the specific use

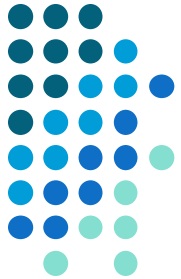
Secondary Data

- Any data collected by someone other than EPA and not under EPA control (states, industries, self-monitoring, etc.)
- Data collected by EPA or others for purpose other than current intended use
- Data compiled from a variety of sources and published in the literature
- Anecdotal information not collected in organized manner
- Models and results from models developed by other organizations or for a different purpose

Existing Data – Specific Examples

- Peer-reviewed journal articles
- Trade magazines
- Maps
- Queries of States, organizations, trade associations, etc.
- Toxicity, exposure, and environmental fate data
- Economic and statistical data not generated specifically for the project
- Census data
- GIS data

Existing Data – Planning Process Questions



- What decision are we trying to make?
- What are the implications of a wrong answer?
- How much information is available?
- What information is directly related to the decision?
- What are the minimum requirements for use of these data?

The 3 most important words in Environmental Decision-making

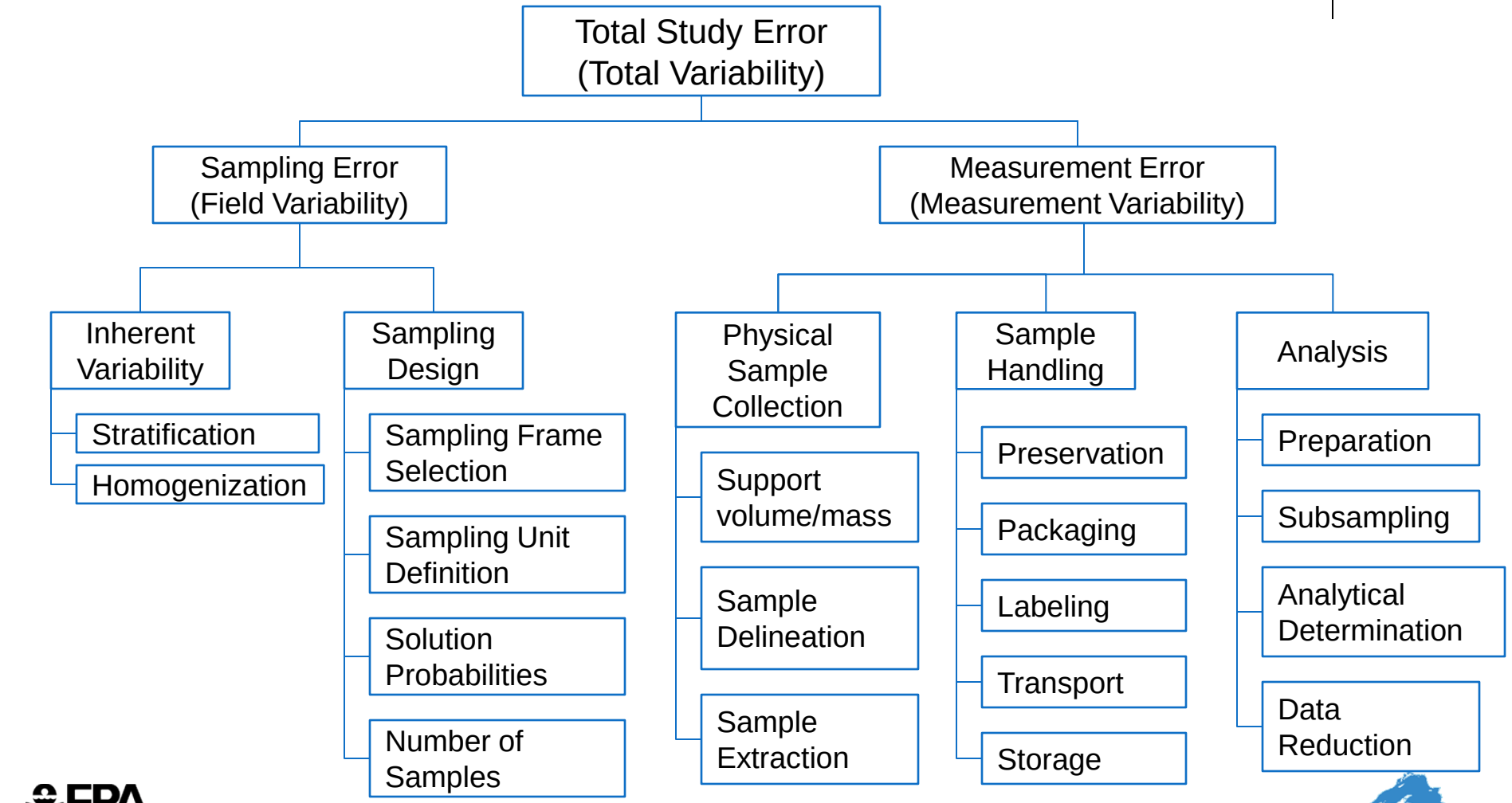
1. Planning!
2. Planning!
3. Planning!

Systematic Planning

- Required for EPA and EPA-funded projects to define performance and acceptance criteria for new and/or existing data
- Guidance available in *Guidance on Systematic Planning using the Data Quality Objective Process,* EPA QA/G-4, EPA/240/B-06/001, February 2006.

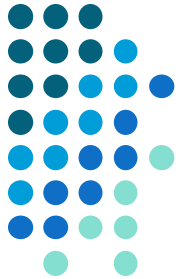
  http://www.epa.gov/quality/qs-docs/g4-final.pdf

Many Sources of Potential Error to Address during Planning



EPA United States Environmental Protection Agency

Estimating Error for Systematic Planning



- Estimating likely sources of error can help determine your sampling strategy and quality objectives
  - If high variability, may need to sample a larger population
  - If larger sample population is beyond resources, you may have to refine your DQOs
- Existing ("secondary") data can help estimate potential error
  - Data generated with the same sampling and analysis methodology for the proposed project
  - Data targeting the same population

Data Quality Indicators and QC Measures

Data Quality Indicator	Meaning	QC Measures
Precision	Agreement among repeated measurements under identical, or substantially similar conditions	▪ Field duplicates or splits ▪ Lab duplicates/replicates * Can be within same organization or among organizations using the same or different methods
Bias	Systematic or persistent distortion of a measurement process that causes errors in one direction	▪ Instrument calibration standards (CAL, VER) ▪ Lab QC spikes (LCS, LFB, OPR) ▪ Matrix spikes & dupes (MS/MSD)

Data Quality Indicators and QC Measures

Data Quality Indicator	Meaning	QC Measures
Accuracy	Overall agreement of a measurement to a known value - includes a combination of random error (precision) and systematic error (bias) in both sampling and analysis operations	▪ Matrix-specific SRMs or CRMs ▪ Spiked matrix samples (MS/MSD, surrogates, isotopically labeled compounds)
Representa-tiveness	The degree to which data accurately and precisely represent a characteristic of a population or condition ☝ *Qualitative DQI– requires BPJ*	▪ No specific QC tools to measure ▪ Evaluate if samples were collected and measurements made in such a way that they reflect the population of interest (as specified in the QAPP)

Data Quality Indicators and QC Measures

Data Quality Indicator	Meaning	QC Measures
Comparability	Measure of confidence that one data set can be compared to another and combined for the decision(s) to be made ☝ *Qualitative DQI – requires BPJ*	▪ Split samples; secondary data ▪ Compare population targeted by sampling techniques; sample collection, handling, preparation, & analysis procedures; holding times, stability issues, QA protocols
Completeness	The amount of valid data needed to be obtained from a measurement system	▪ # of valid results vs. the number determined to be necessary during project planning (as specified in the QAPP)

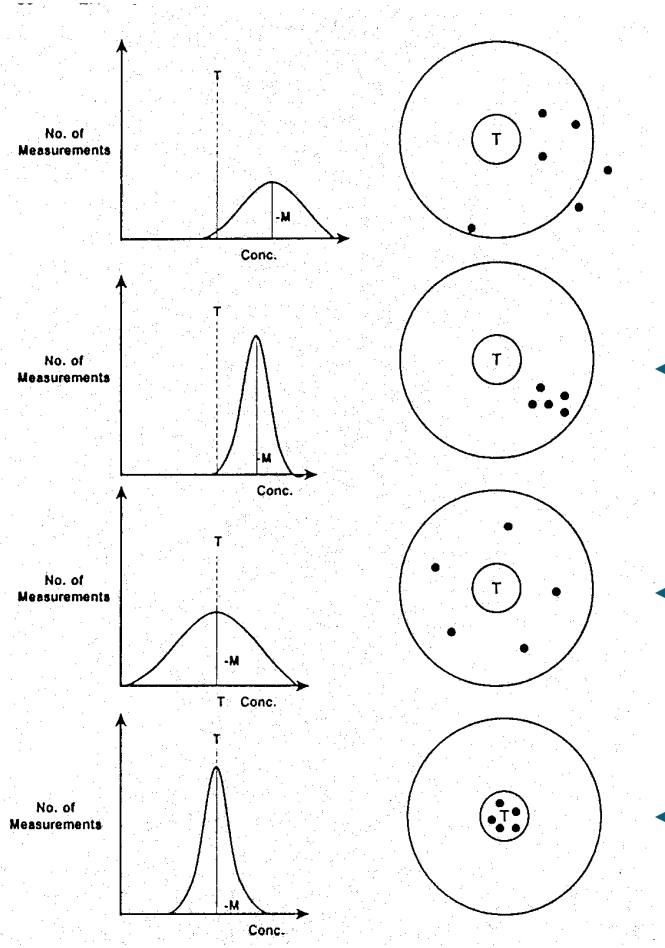
Data Quality Indicators and QC Measures

Data Quality Indicator	Meaning	QC Measures
Specificity	Correct identification of the parameter you are targeting	▪ Retention times ▪ Ion abundance ratios ▪ Confirmation analyses (e.g., alternate GC column) ▪ Peak shape
Detection and Quantitation	The ability to ▪ Determine if it is there or not ▪ Distinguish between responses representing different concentrations of interest	▪ Method Detection Limit (MDL) or equivalent ▪ Signal to noise ratios ▪ Calibration range ▪ Analysis of samples at/near quantitation limit ▪ Well below action level

DQI Impacts

- Precision: decision must allow for error from all sources (sampling, analysis)
  - Best indicator would be replicate sampling and analysis (e.g., field dupes and lab dupes)
- Accuracy: decision must allow for bias in results
  - Recovery correction may be necessary
- Detection and quantitation level: must be well below action level
  - “ND” at or above action level precludes decision

Precision and Accuracy Illustrated



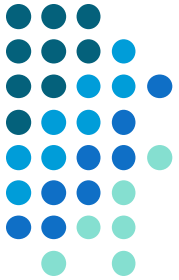
Imprecise and biased. The mean is not close to the true value and the individual results are scattered

Precise, but biased. The mean is not close to the true value but the results are clustered

Unbiased, but imprecise. The mean reflects the true value, but the results are scattered

Precise and unbiased = Accurate

Sampling Design Considerations during Project Planning



- Hot spot(s)
  - If known, sample to attempt to obtain maximum concentrations; if unknown, use random sampling
- Random
  - Area to be covered
  - Representative number of samples
  - Involve statistician
    - Get sampling error for spatial distribution – objective is to quantify error components
- Contamination control (*example on next slide*)

Guidance for Choosing a Sampling Design for Environmental Data Collection, December 2002 http://www.epa.gov/quality/qs-docs/g5s-final.pdf

EPA United States Environmental Protection Agency

Planning for Contamination Control Example – Mercury

- Mercury is one of the most difficult substances to collect and measure
  - Ubiquitous – it's everywhere
    - Great Lakes POTW study – the Hg blank collected on the only rainy day had a Hg concentration 3x higher than any other field blank in the study even with "clean" sampling techniques
  - Mercury is volatile – it moves around
  - Toxic – measurement at very low levels required
    - Water quality criterion of 1.3 ppt (ng/L) for protection of wildlife in the Great Lakes States and Tribes
  - Amalgamates with other metals
- Affects both sampling and analysis strategies

Planning for Contamination Control Example – Mercury (cont'd)

- Contamination must be eliminated or reduced to a level that will not compromise the measurement
  - Sample bottles (sampling)
  - Sampling equipment (sampling)
  - Sampling procedures-"clean hands/dirty hands" (sampling)
  - Reagents (sampling and analysis)
  - Laboratory environment (analysis)
  - Laboratory glassware and equipment (analysis)
- Need to address this challenge during planning
  - Good data depends on it

Planning for Contamination Control Example – Mercury (cont'd)

- Plan and budget for appropriate sampling and analysis strategies to ensure unbiased data
  - Enhanced equipment cleaning
  - Equipment blanks to verify
  - 2-person (clean hands/dirty hands) sampling
  - Consider contamination sources in the field (wind & stream direction, metal supports, bridges, roads)
  - Extra field and laboratory blanks to monitor, control, and document contamination

Analysis Considerations in Planning

Choose the right method!

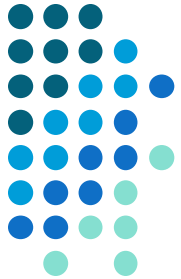
- Will it work in your matrices? Matrix problems can prevent
  - Recovery of pollutants at normal levels
  - Achievement of quantitation levels needed to support your decision
  - Example: Drinking water methods don't always work on wastewaters or ambient waters with high sediments
- Tips/techniques for matrix challenges available in EPA's "Pumpkin Book"

  Reference: *Solutions to Analytical Chemistry Problems with Clean Water Act Methods*, March 2007. http://water.epa.gov/scitech/swguidance/methods/atp/upload/2008_02_06_methods_pumpkin.pdf

Analysis Considerations in Planning – Choosing the Right Method

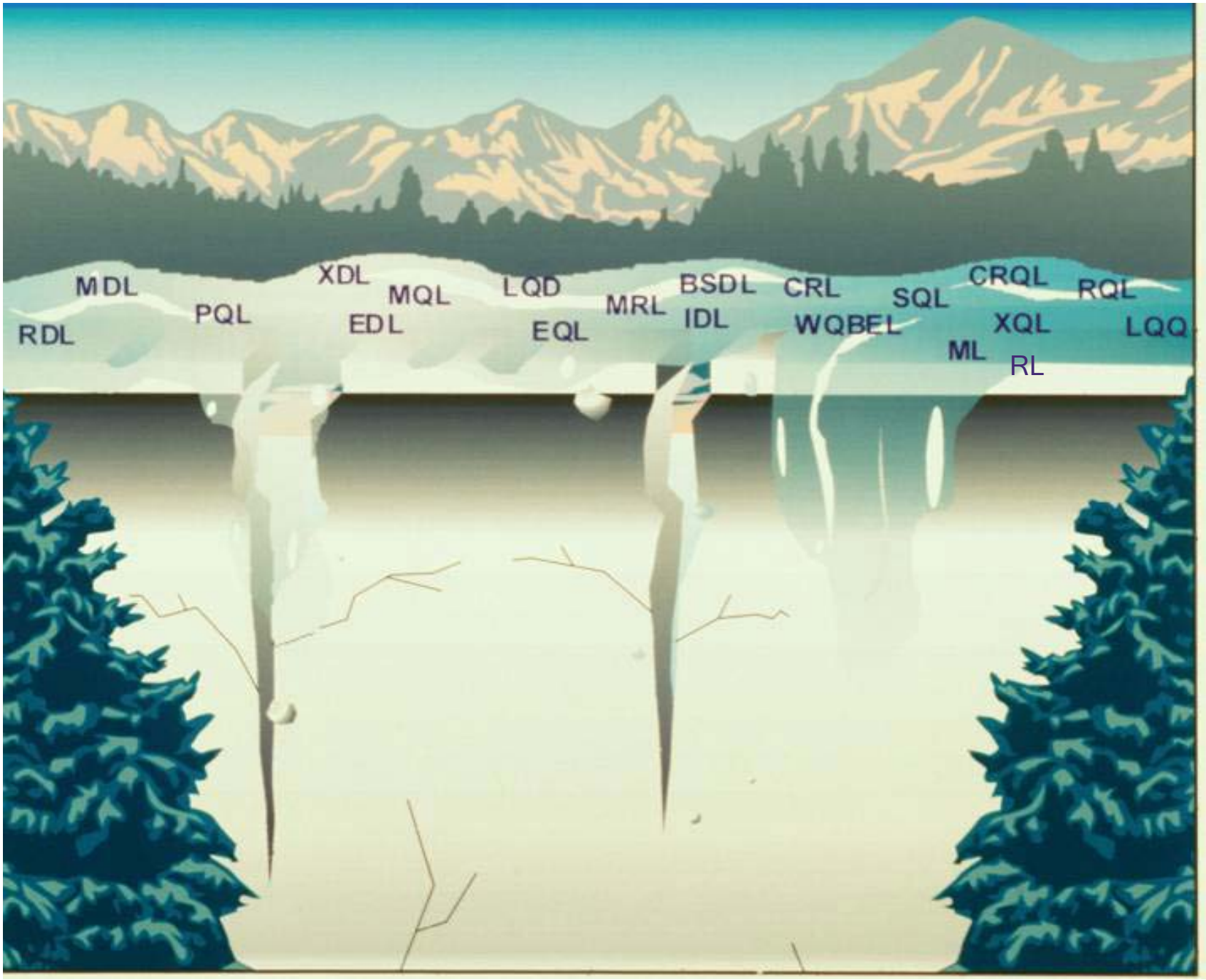
- Does the method have QC elements and acceptance criteria that support project DQOs?
  - If not you may need to add/specify them
- Does it support applicable regulatory requirements?
  - e.g., CWA and SDWA compliance monitoring both require use of *approved* methods (40 CFR parts 136 and 141)
- Is it comparable to historical data?
  - Ideally yes, unless you deliberately need something different
- Does it achieve the detection and quantitation limits you need?
  - DL 10x < your action level (ideally 20x)
  - QL no higher than your action level
- Is it overkill?
  - Does it measure pollutants not pertinent to the decision?
    - Many methods are for broad-range screening
    - A method optimized for a specific pollutant may provide more reliable data

Analysis Considerations in Planning *(continued)*



Detection, Quantitation, Reporting Limits

- Understand the terms used and how they relate to your needs
  - More than 50 different terms have been used and there is no consensus on best approach
- Generally speaking
  - **Detection limits** are the lowest concentration that allows for differentiation between a sample that contains a substance and one that does not (is it there or not?)
  - **Quantitation Limits** are the lowest concentration that can be measured with some degree of confidence (can I actually tell how much is there?)
  - **Reporting Limits** can be either of the above... Be careful!


RDL
MDL
PQL
XDL
EDL
MQL
LQD
EQL
MRL
BSDL
IDL
CRL
WQBEL
SQL
CRQL
ML
XQL
RL
RQL
LQQ

Planning/Estimating Analysis Error – Example

Cleanup level of 10 (pick your units) required

- Analytical method measures with 20% low bias (average 80% recovery) and 30% RSD error in precision
  - With 20% bias, result could be 8; including 30% precision error, result could be as low as 5.6. To allow for this error, cleanup to 5.6 (not 10) may be required.
- Doesn't include sampling error
- If decision is critical, isotope dilution should be considered. Bias is typically zero, and RSD is one-half that of internal standard methods

Lab Considerations

Choose the right lab

- Does it have the capability and capacity to meet your schedule?
- Does it have experience with
  - The analytical method?
  - The matrix?
  - Your reporting requirements?
- Can it prove the above?
  - Accreditations are only one of many tools to evaluate capability. Others may be:
    - Historical data using same method/matrices, control charts
    - SRMs/CRMs (in your matrix), MDL studies, IPR studies
    - Lab QA program (SOPs, audit results)

Lab Considerations – Lab QC

Some elements of lab QC

- Quality System in place – QMP or QA Manual
- Purity & traceability of standards and reference materials
- Calibration procedure, range, linearity, and verification
- Detection, quantitation, and reporting level
- Frequency of blanks and lab control samples
- Spikes and duplicates for each separate sample matrix (not required for isotope dilution methods)
- Field duplicates to quantify sampling uncertainty
- Statements of data quality or QC charts (not required but helpful)

Data Reporting Considerations

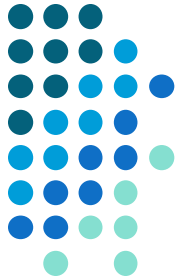
Reporting is critical

- Can't overlook it during planning phase
- How will you handle results below detection or quantitation limits?
- What data elements do you need?
  - Just the field results?
  - Supporting QC results?
  - Raw instrument data and log notebooks?

Data Reporting Considerations (cont.)

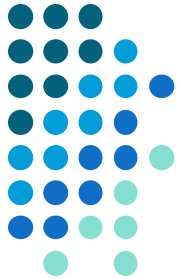
- How do you need the data reported?
  - Hardcopy vs. electronic?
  - Standardize format or let the labs decide?
  - Will the labs apply qualifiers (flags) or will you? Or both?
    - May depend on
      - Your data review approach
      - Size/cost of project
      - Size of supporting labs

Data Reporting Formats- Hardcopy Options



- Easy to maintain integrity (can't modify hardcopy without marking it up)
- Requires manual data review
- Customized, standardized forms that summarize data
  - Pros:
    - Standard, summary level format
    - Easy to read and interpret
  - Cons:
    - If not well designed, hard to export from instruments or Laboratory Information Management Systems (LIMS)
    - Requires data entry to electronically review or manipulate data
- Format of lab's choice
  - Pro: May decrease costs for lab
  - Cons:
    - Still requires data entry for electronic data review/manipulation
    - May increase costs for data review

Data Reporting Formats – Electronic Options



- Harder to protect integrity of original data (usually easy to modify the file, requires strong ‘version control’)
- Can range from simple, lab-designed spreadsheets to highly complex electronic data reporting systems (e.g., CLP SEDD) to things in between (e.g., EPA OW EAD)

Data Reporting Formats – Standardized Nomenclature

- Consider need to standardize data elements (not just format)
  - Analyte name (e.g., PCB-114 vs. 2,3,4,4',5-PeCB)
  - Qualifiers (one lab may use "<" to indicate result below reporting threshold, another may use "U")
  - QC sample names (e.g., LFB vs. LCS vs. OPR)
- Applies to both electronic and hardcopy data

Data Reporting- Lab Cover Page

- Labs should provide a cover page that
  - Discusses (not just mentions!) any issues associated with the data
  - Defines all qualifiers (flags) that the lab applied to the data
  - Contains a signature of the lab manager certifying the analytical results.
- Applies to both hardcopy and electronically reported data

Example from a Lab Cover Page

The response of co-eluting PCBs 61/70/74/76 in 61512 ID L4815-14) was above the calibrated linear range on the instrument. As contributions to the response were from multiple congeners, and the response of the contributing congeners would each be within the calibrated linear range the determined concentration was judged to be accurate. No further work was performed.

Analysis batch CLWG8932

Instrumental re-analysis of samples listed below was performed to confirm the possible contribution to the response of some analytes from a high level sample analyzed just prior. The results from both analyses were in good agreement for each sample; results from the initial analysis data are reported.

Client Sample ID	Sample ID
61624	L5047-10
61652	L5143-3
63005	L5257-20

DATA PACKAGE

Included in the data package are the narrative, a list of modifications of the USEPA I… Instrumental gas chromatograph temperature programs, sample Cover Page and Corre…on …le, laboratory extraction logs, sample data reports, laboratory QC data reports and instrument QC.

I certify that this data package is in compliance with the terms and conditions of the contract both technically and for completeness, for other than the conditions detailed above, …dditio… certify, that to the best of my knowledge and belief, the data as reported a… …ccurate. Release of the data contained in this data package has been authorized by t… …ory Manager or her designee, as verified by the following signature.

Signed: M.Sc. Project Chemist

19 Jun … Date Signed

Data Reporting Forms

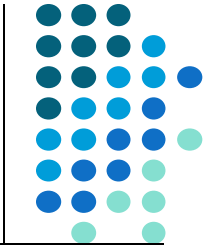
- Hardcopy or electronic summary of analytical results for the samples collected
  - Summary may also contain results of supporting QC
  - CLP-style reporting forms are ‘summary level’
  - Can also be in spreadsheet format
- Should include:
  - Sample number (ID)
  - Analyte
  - Result and/or detection limit
  - Units
  - Analysis data and time
  - Lab qualifiers

Example of Summary Level Lab Data (Hardcopy Reporting Form)

Form 1A
VOLATILE ORGANICS ANALYSIS DATA SHEET – TARGET ANALYTES
Use for Sample and Blank Results

Lab Name:	Contract No:			EPA Sample No.	
Analytical Method:	Episode No:			Lab Sample ID:	
Matrix (aqueous/solid/tissue):				Lab File ID:	
Date Received:				Instrument ID:	
Sample Wt or Vol:	g or mL:			Column:	
% Solid/Lipid:	Analysis Date:			Analysis Time:	
Dilution Factor:		Concentration Units (ug/L or ug/kg):			
Analyte		CAS No.	Concentration Found	Q	Reporting Limit (ML)

Example of Summary Level Lab Data (Excel Spreadsheet)



Microsoft Excel - CL_1111.xls

File Edit View Insert Format Tools Data Window Help

MS Sans Serif 10

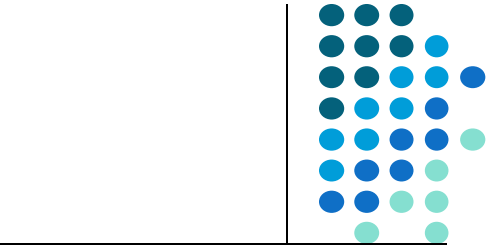
A12 =

	A	B	C	D	E	F	G	H	I	J	K
1	AMOUNT	DETLIMIT	ANALYTE	METHOD	UNIT	SAMPLE	PREPDATE	ANALDATE	ANALHOUR	ANALMIN	DILUTION
2	0.1	1	AMMONIA AS NITROGEN	350.2	MG/L	99999	6/24/1998	6/24/1998	17	34	1
3	0	1	AMMONIA AS NITROGEN	350.2	MG/L	99998	6/24/1998	6/24/1998	17	36	1
4	314	6	AMMONIA AS NITROGEN	350.2	MG/L	99997	6/24/1998	6/24/1998	17	40	1
5	319	6	AMMONIA AS NITROGEN	350.2	MG/L	99996	6/24/1998	6/24/1998	17	46	1
6	0.89	1	AMMONIA AS NITROGEN	350.2	MG/L	99995	6/24/1998	6/24/1998	17	59	1
7	0	1	AMMONIA AS NITROGEN	350.2	MG/L	99994	6/24/1998	6/24/1998	19	54	1
8	0.08	1	AMMONIA AS NITROGEN	350.2	MG/L	99993	6/24/1998	6/24/1998	20	4	1
9	0.15	1	AMMONIA AS NITROGEN	350.2	MG/L	99992	6/24/1998	6/24/1998	20	9	1
10	0.39	1	AMMONIA AS NITROGEN	350.2	MG/L	99991	6/24/1998	6/24/1998	20	17	1
11											
12											
13											

Additional Data Needed for Validation

- Summary level QC data
- Raw data (bench sheets, calculations, instrument printouts, etc.)
- Analytical run chronology (date & time)
- Sample receipt log
- Specific electronic reporting format that will help automate the data review and/or data management processes

Example Instrument Printout



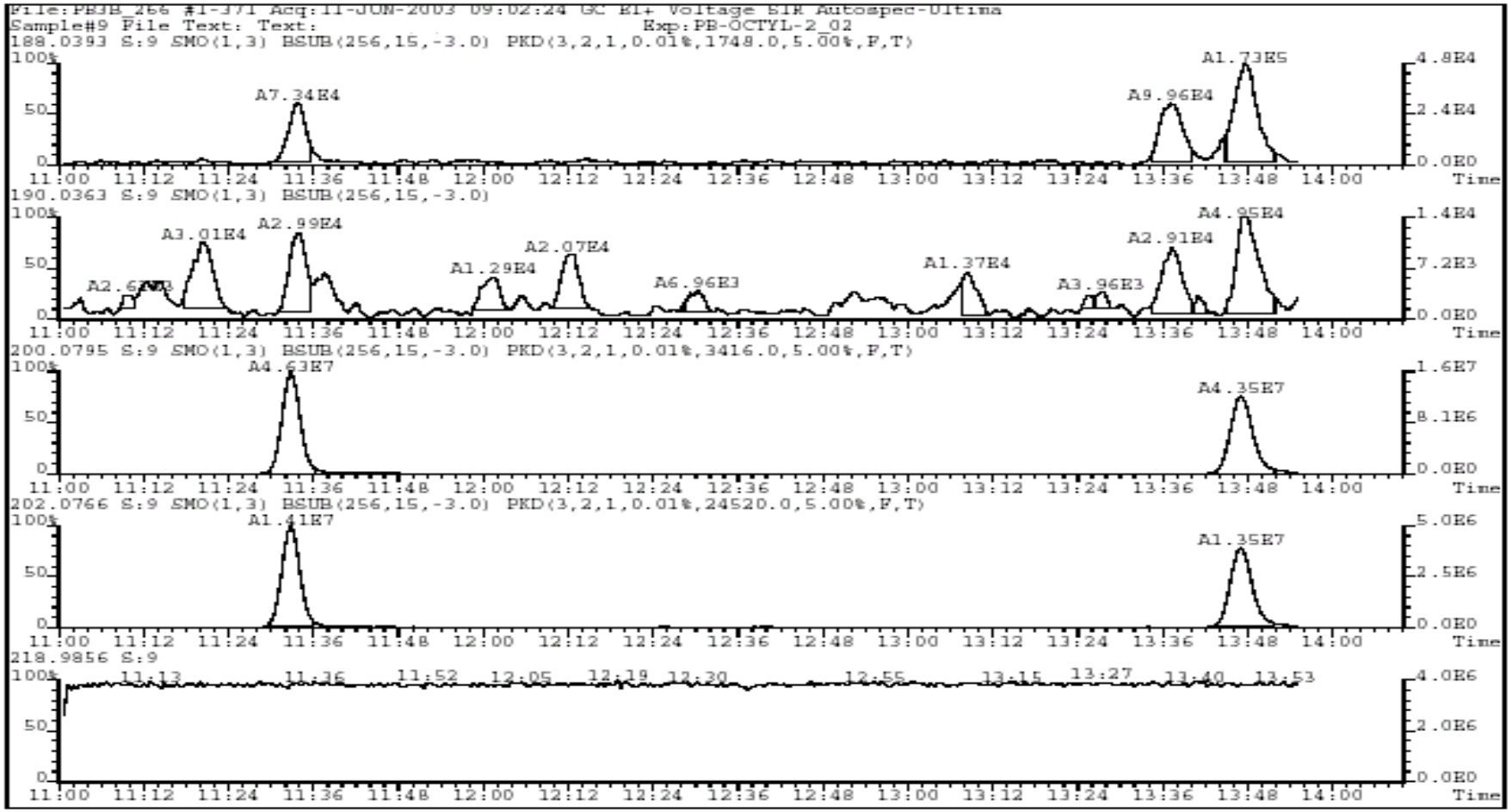
OPUSquan 11-JUN-2003 Page 9

Page 3 of 3

Run #4 Filename PB3B_266 S: 9 I: 1 Acquired: 11-JUN-03 09:02:24 Processed: 11-JUN-03 10:40:54
Run: pb3b_266-» Analyte: 1668xa-s3 Cal: pb3b_266-» Results: pb3b_266-» Version: V3.6 6-JAN-2000 17:51:42
Sample text: L4926-6,I,61546 Comments: 1,WG8864,1.0/22uL
sample size: 8.134300 conc units: pg/g total toxicity: 6228.41 F1: 1.0000 F2: 1.0000

	Typ	Name	#Hom	Resp	RA		RT	Conc	Tox #1	DL	Rec	M?
1	Unk	CL1-PCB-2	1	1.29e+05	3.42	y	13:37	0.624	1	0.1468	-	y
2	Unk	CL2-PCB-10	0	*	*	n	NotFnd	*	*	0.2256	-	n
3	Unk	CL2-PCB-9	1	3.87e+04	1.07	n	16:06	0.142	0	0.2184	-	y
4	Unk	CL2-PCB-7	1	3.14e+05	1.68	y	16:16	1.150	1	0.2186	-	y
5	Unk	CL2-PCB-6	1	5.33e+04	1.82	n	16:30	0.197	0	0.2206	-	y
6	Unk	CL2-PCB-5	0	*	*	n	NotFnd	*	*	0.2350	-	n
7	Unk	CL2-PCB-8	1	3.54e+05	1.51	y	16:58	1.233	1	0.2078	-	y
8	Unk	CL2-PCB-14	0	*	*	n	NotFnd	*	*	0.2344	-	n
9	Unk	CL2-PCB-11	1	4.27e+05	1.53	y	19:31	1.869	2	0.2613	-	y
10	Unk	CL2-PCB-12/13	0	*	*	n	NotFnd	*	*	0.2598	-	n
11	Unk	CL3-PCB-30/18	1	6.41e+05	1.04	y	19:11	2.211	2	0.0520	-	n
12	Unk	CL3-PCB-17	1	2.82e+05	1.08	y	19:36	1.225	1	0.0653	-	y
13	Unk	CL3-PCB-27	1	6.52e+04	1.19	y	19:50	0.196	0	0.0453	-	y
14	Unk	CL3-PCB-24	1	3.13e+04	1.28	n	19:58	0.093	0	0.0446	-	y
15	Unk	CL3-PCB-16	1	1.30e+05	0.97	y	20:05	0.638	1	0.0738	-	y
16	Unk	CL3-PCB-32	1	1.42e+05	1.16	y	20:37	0.508	1	0.0758	-	y
17	Unk	CL3-PCB-34	0	*	*	n	NotFnd	*	*	0.0789	-	n
18	Unk	CL3-PCB-23	0	*	*	n	NotFnd	*	*	0.0801	-	n
19	Unk	CL3-PCB-26/29	1	2.35e+05	0.99	y	22:24	0.892	1	0.0805	-	y
20	Unk	CL3-PCB-25	1	9.33e+04	0.90	y	22:40	0.307	0	0.0696	-	y
21	Unk	CL3-PCB-31	1	2.44e+06	1.03	y	22:58	8.829	9	0.0767	-	y
22	Unk	CL3-PCB-28/20	1	3.13e+06	0.98	y	23:17	11.522	12	0.0780	-	n
23	Unk	CL3-PCB-21/33	1	2.76e+05	0.96	y	23:32	0.960	1	0.0738	-	y
24	Unk	CL3-PCB-22	1	2.61e+05	0.88	y	23:56	0.999	1	0.0809	-	y
25	Unk	CL3-PCB-36	0	*	*	n	NotFnd	*	*	0.0801	-	n
26	Unk	CL3-PCB-39	1	3.10e+04	0.76	n	25:59	0.112	0	0.0763	-	y
27	Unk	CL3-PCB-38	0	*	*	n	NotFnd	*	*	0.0844	-	n
28	Unk	CL3-PCB-35	0	*	*	n	NotFnd	*	*	0.0895	-	n
29	Unk	CL4-PCB-50/53	1	2.08e+05	0.73	y	22:40	0.766	1	0.0165	-	n
30	Unk	CL4-PCB-45/51	1	9.13e+05	0.77	y	23:26	3.488	3	0.0172	-	n
31	Unk	CL4-PCB-46	1	3.27e+04	0.65	n	23:40	0.151	0	0.0207	-	y
32	Unk	CL4-PCB-52	1	8.57e+06	0.78	y	25:10	31.638	32	0.0166	-	y
33	Unk	CL4-PCB-73	0	*	*	n	NotFnd	*	*	0.0130	-	n
34	Unk	CL4-PCB-43	1	5.61e+04	0.74	y	25:25	0.263	0	0.0211	-	y
35	Unk	CL4-PCB-69/49	1	4.87e+06	0.79	y	25:41	16.462	16	0.0152	-	n
36	Unk	CL4-PCB-48	1	5.06e+05	0.84	y	25:59	1.986	2	0.0176	-	y
37	Unk	CL4-PCB-44/47/65	1	8.03e+06	0.78	y	26:15	29.432	29	0.0165	-	n
38	Unk	CL4-PCB-59/62/75	1	6.72e+05	0.75	y	26:34	2.036	2	0.0136	-	y
39	Unk	CL4-PCB-42	1	4.99e+05	0.80	y	26:46	2.127	2	0.0191	-	n
40	Unk	CL4-PCB-41/40/71	1	6.38e+05	0.73	y	27:15	2.569	3	0.0181	-	y
41	Unk	CL4-PCB-64	1	5.74e+06	0.80	y	27:31	16.794	17	0.0132	-	n
42	Unk	CL4-PCB-72	1	3.13e+05	0.69	y	28:23	1.019	1	0.1998	-	y
43	Unk	CL4-PCB-68	1	1.23e+06	0.71	y	28:41	4.000	4	0.2000	-	y
44	Unk	CL4-PCB-57	1	5.30e+04	0.84	y	29:07	0.181	0	0.2098	-	y

Example Chromatogram



Example Run Chronology

Experiment : PB-DB1-1_01
GC Program : PB-DB1-1_01
Column type : DB1
Serial # : 2184312H
kPa : 150
Vol injected: 2.0
PMT Voltage : 371

Temps -source: 270 Tune : GW
-s resv: 160 List : TD
-re_ent: 280 Check :
-cap_1 : 280 LIMS :
-cap_2 : 280

Date -list : 30-MAY-2003
-liner : 30-MAY-03 R,RES
-septum: 30-MAY-2003
-guard : 40cm 30-MAY
-column: NEW 19-FEB-03
-t line: 2cm 31-MAR-03
-bake : 30hrs 17-MAY-03
-source: 10-MAR-2003

#	Data file	S	V	Sample Text	Comments	Acquisition Date/Time
1	PB33_163A	1	1	PB067A-CAL,,/1-01	1,,2.0uL CS-0	30-MAY-03 22:58:25
2	PB33_163A	2	2	PB067B-CAL,,/1-01	1,,2.0uL CS-1	30-MAY-03 23:38:03
3	PB33_163A	3	3	PB067F-CAL,,/1-01	1,,2.0uL CS-5	31-MAY-03 00:17:41
4	PB33_163A	4	4	PB067F-CAL,,/1-01	1,,2.0uL CS-5	31-MAY-03 00:57:22
5	PB33_163A	5	5	PB067E-CAL,,/1-01	1,,2.0uL CS-4	31-MAY-03 01:36:59
6	PB33_163A	6	6	PB067D-CAL,,/1-01	1,,2.0uL CS-3	31-MAY-03 02:16:38
7	PB33_163A	7	7	PB067C-CAL,,/1-01	1,,2.0uL CS-2	31-MAY-03 02:56:15
8	PB33_163A	8	8	TOLUENE,,	1,,2.0uL	31-MAY-03 03:35:55
9	PB33_163A	9	9	LS395-20,,63037	1,WG8838,2.0/22uL	31-MAY-03 04:15:37
10	PB33_163A	10	10	LS395-22,,63039	1,WG8838,2.0/22uL	31-MAY-03 04:55:14
11	PB33_163A	11	11	LS395-24,,63043	1,WG8838,2.0/22uL	31-MAY-03 05:34:51
12	PB33_163A	12	12	LS649-2,,63055	1,WG8838,2.0/22uL	31-MAY-03 06:14:27
13	PB33_163A	13	13	LS649-4,,63057	1,WG8838,2.0/22uL	31-MAY-03 06:54:03
14	PB33_163A	14	14	LS649-6,,63059	1,WG8838,2.0/22uL	31-MAY-03 07:33:39
15	PB33_163A	15	15	LS649-8,,63063	1,WG8838,2.0/22uL	31-MAY-03 08:13:16
16	PB33_163A	16	16	LS649-10,,63065	1,WG8838,2.0/22uL	31-MAY-03 08:52:53
17	PB33_163A	17	17	LS649-12,,63067	1,WG8838,2.0/22uL	31-MAY-03 09:32:28
18	PB33_163A	18	18	PB067D-CAL,,/1-01	1,,2.0uL CAL VER	31-MAY-03 10:12:04

Example Lab Receipt Log

PCB CONGENERS BY EPA METHOD 1668A
PBDE/PBB ANALYSES –

PCBs: 1668A MLA-010 Rev 4 ✓ PBDEs: MLA-025 Rev 2 _____ PBBs: MLA-026 Rev 1 _____
Sample ID: 14815-3 Date: 27 May-03 Analyst: _____ BATCH # CLLib 8859
Blank: WB 8859-lb1 Rec'd: 9 Jul-02 Back-up: ~ 60g Stored In: WIF 4 Date Due: ASAP.
Sample Type and Description: greyish white, chunky tissue with very small black and white flecks.
Sample Pre-treatment: Homo: 6 Feb-2003
Original Labelling: EPA No. 1 of 1 1601

Sample Weight: Wet 10.07 g % Moisture _____ Dry _____ g
Tare: _____ g
Dry + Tare: _____ g Date: _____ Time: _____
Wet + Tare: _____ g Date: _____ Time: _____
Dry/Wet Ratio: _____ Balance Inventory #: 3043

AUTHENTIC STANDARD

name	quantity	date	time	analyst	witness

SURROGATE STANDARD

name	quantity	date	time	analyst	witness
PBO15A-Sur/01	20 mL	27May03	14:46		YMS

DRYING: Section 5.1.1 Weight Na_2SO_4 ~ 100g Drying Time ~ 1hr. Analyst _____
EXTRACTION: Section 5.1.2
1. Load soxhlet thimble: _____ 3. 300 mL DCM + boiling chips: _____
2. Add Surrogate Standard to thimble: ✓ 4. Extraction: Begin ~~1145~~ End 0800

CLEAN-UP STANDARD

name	quantity	date	time	analyst
PBO15A-SUR/16	20uL	28May03	1645 1107	AZJ

Rotovap to 1 mL If no lipid analysis: ✓ to 5 mL if lipid analysis: _____
Lipid Analysis: Avg. % Lipid _____ Recovery Correction Factor _____ Analyst _____ *(see reverse for details)*
Reduce volume. Add 1 mL DCM.
Section 5.1.3 SX3 Biobeads: Discard 140 mL Collect 160 mL Analyst GZJ Column No. 50
Reduce Volume, transfer with Hexane ✓
CLEANUP Section 6.1 Acid/Base Silica: ✓ copper ✓ negative
Section 6.2 Florisil: Batch No. 283 45 (mL, E1) copper ✓ neg.

Staged Electronic Data Deliverable (SEDD)

- What is SEDD?
  - Hierarchal file of results for a data package (“SDG” for CLP)
    - Created by a laboratory information management system (LIMS)
    - Transmitted to EPA and data reviewers electronically
    - Allows automated review of data in SDG
- SEDD benefits
  - Uniform electronic format for data
  - Standardized definition of records for methods, analytes, units, etc.
  - Allows delivery of information for automated review of data at various levels
    - Comparing results to specifications
    - Recalculation of results
- SEDD limitations
  - Costly initial investment. May not fit all situations:
    - Small data packages
    - Non-routine analyses

SEDD Stages

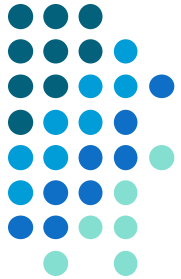
- Stage 1 – Sample results only (e.g., CLP Form I)
- Stage 2a – Sample results and method QC data
- Stage 2b – Sample results and method and instrument QC results
- Stage 3 – Stage 2B plus raw results enabling independent sample recalculation
- Stage 4 – Raw instrument data files (GC/MS chromatograms & spectra); currently under development

http://www.epa.gov/superfund/programs/clp/sedd-geninfo.htm#geninfo

EPA OW EAD Approach

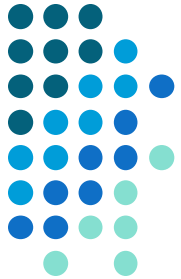
- Goal
  - Standardize reporting to facilitate consistency
  - Shift to automated reporting streamline data review and data management
  - Avoid extensive cost burden on lab
- Each lab must submit
  - Sample receipt acknowledgement (e.g., signed traffic reports)
  - MDL study data for applicable method/matrix
  - Initial precision & recovery (IPR) study data for applicable method/matrix
  - Complete sample data package
    - Combination of hardcopy and electronic format requirements *(next 2 slides)*

EPA OW EAD Approach – Hardcopy Data Requirements



- **Data narrative**
  - Summarizes what's in the package (lab name, project/contract #s, sample IDs...)
  - Discussion of issues with samples, shipment, or analysis (including QC failures)
  - Lab Manager Certification of results submitted (signature)
- **Summary data result forms for all field and QC samples**
  - Standard form (similar to CLP), but flexibility is allowed for alternative versions that capture each data element
  - Alternative requires pre-approval
- **Raw data**
  - Raw instrument or data system printouts, logbook pages and manual worksheets, bench sheets, chromatograms, etc.
  - Also includes standard mix prep documentation logbook pages for all stock and working standard mixes
- **Supporting documentation (e.g., Traffic Reports)**

EPA OW EAD Approach – EDD Requirements

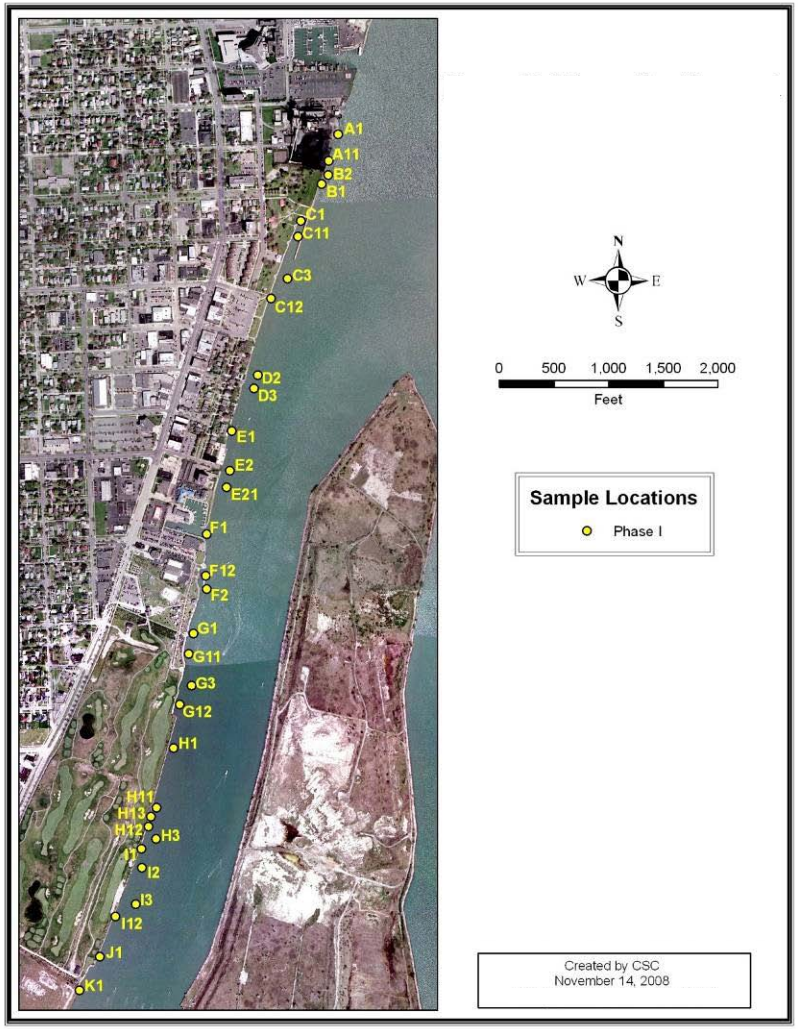


- 3 separate, comma-delimited ASCII text files or Excel CSV files
  - **Analytical Results File (A1)** – contains analytical results and related info on an analyte level for field samples and associated lab QC (excluding calibrations and instrument tunes)
  - **Laboratory Instrument File (A2)** – contains results and related info for initial calibration standards and calibration verifications
  - **Sample Analysis File (A3)** – info on sample level for field samples and lab QC samples (excluding calibrations and tunes)
  - ***Sound familiar?*** Structure designed to be compatible with SEDD Level 2B and with commercial ADR software
- Standardized field names and field nomenclature
  - Ensures consistency in analyte names, QC element types, method names, etc.

Tying it all together

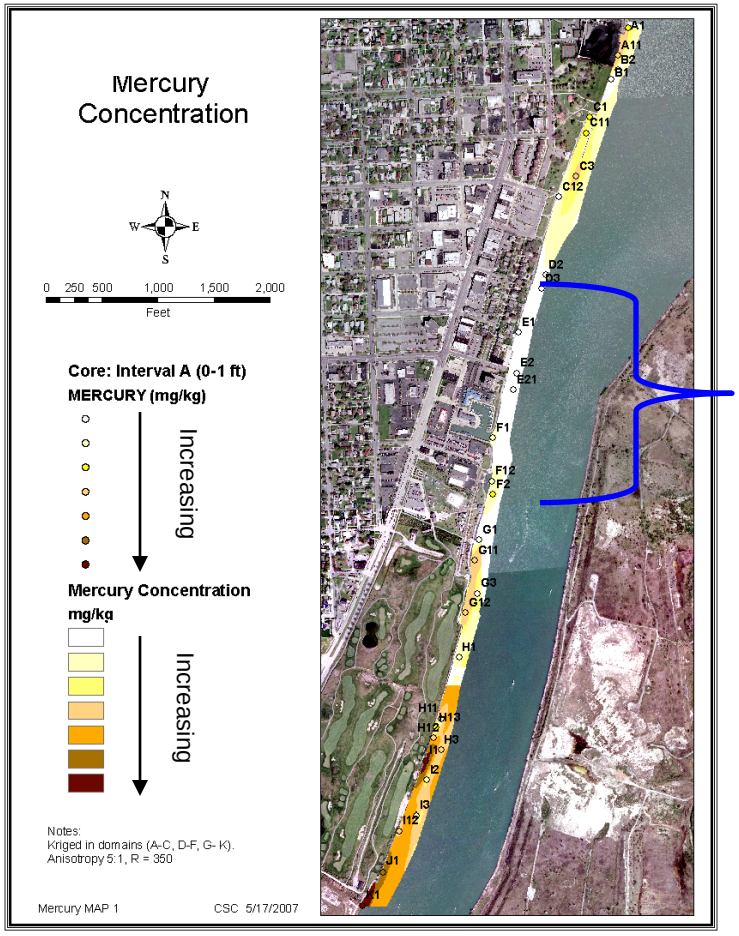
- Data Quality Act & Info Quality Guidelines require us to maximize and document the quality, objectivity, utility and integrity of data/info generated with EPA $$
- EPA's Quality System supports this
  - Requires systematic planning to identify data needs, intended us, and procedures to control, evaluate and document quality
  - Allows use of existing data as long as it of acceptable quality to support project planning needs and/or project decisions
  - Requires a QAPP to document the systematic planning results
- Examples (next few slides)

Planning Example: Remedial Investigation Project



- Questions based on review of Phase 1 results
  - Is there an increasing trend in Hg concentrations as you move north from Transect F to A?
  - What is the distribution of Total PCBs (as Aroclors) within Transects B and C?
  - Are the contaminant concentrations in transects D, E, and F below thresholds of concern (TOCs)

Planning Example: Remedial Investigation Project (cont'd)



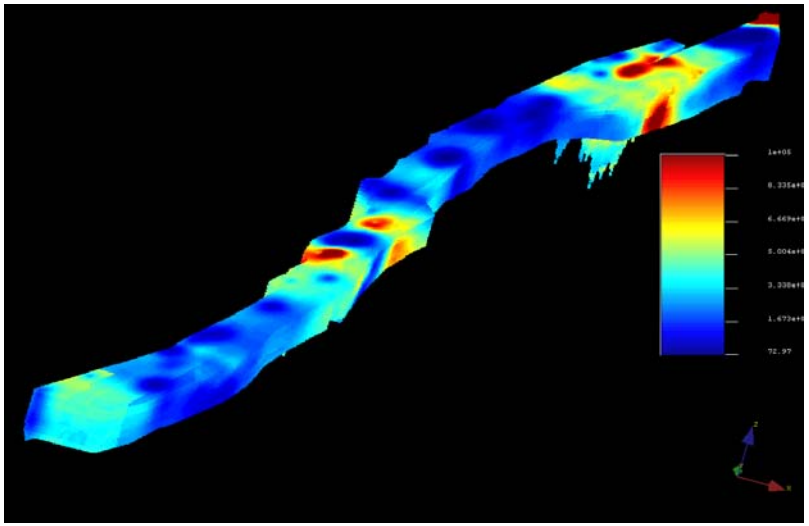
Distinct Populations

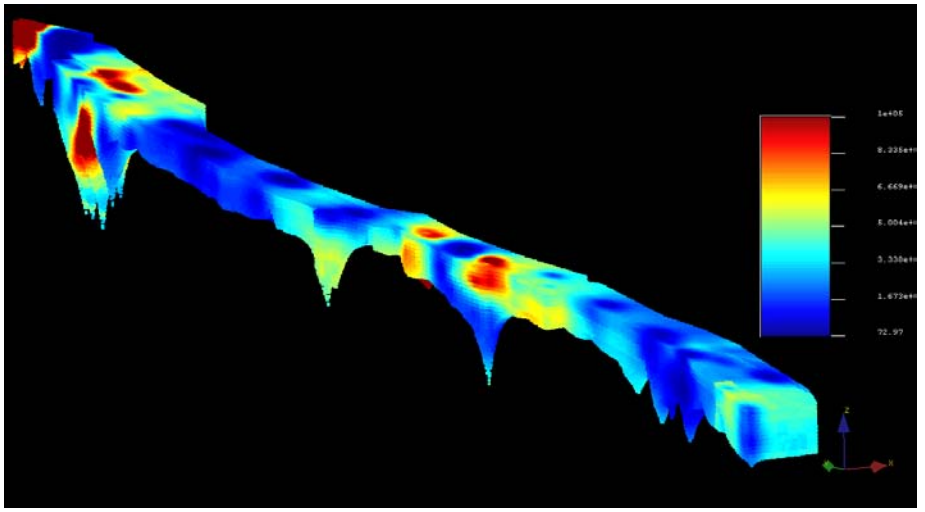
- Conducted statistical and geostatistical analysis of the Phase 1 data to generate the sampling design
- First step: Understand the data!
  - Data comparability
    - Site boundaries, recent data, same compounds of interest, same sampling procedures
  - Exploratory data analysis
    - Investigate populations
    - Evaluate the distribution

Planning Example: Remedial Investigation Project (cont'd)

- Used systematic planning to develop a statistical sampling design to answer each specific question
  - Outlined decision statement, sampling design, and intended data interpretation procedure
- Final data set from Phase 1 and Phase 2 were used to model contaminant concentrations across the site (next slide)
  - Supports remediation planning activities

Planning Example: Remedial Investigation Project (cont'd)





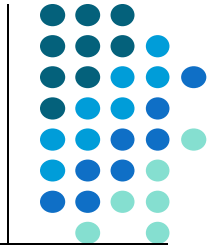
Reporting Example – National EPA Biosolids Assessment Study

- Large multi-year EPA study involving
  - Multiple EPA Program offices, and 3 EPA labs
  - State partners
  - Four support contractors and one lab contractor
  - Main study and focused sub-studies
- In-depth planning among offices→ 2 project QAPPs
  - One covered all sample prep & all EPA lab analyses
  - One covered contractor lab analyses
- EPA lab QAPP specified all data elements to be reported and that the EPA labs would use a standard data reporting format.

Reporting Example – National EPA Assessment Study *(cont.)*

- How did it turn out?
  - Each operational lab (e.g., the organics lab vs. the metals lab) within the same EPA lab facility:
    - Had standardized, but different reporting formats
    - Contained most, but not all the different critical elements needed
    - *Examples on the next 2 slides*
  - Each 'standardized format' required significant manipulation to:
    - Review the data in an efficient, organized way
    - Manipulate the data for inclusion in the study database
- Lesson:
  - Specify the 'standardized' reporting format or review the lab's proposed reporting format ahead of time if data review/data management resources are limited

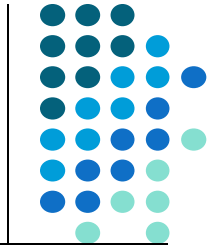
Reporting Example (cont.) – Hg Data Format



	A	B	C	D	E	F	G	H
1	*[Project Name]*							
2	**Mercury Analysis**							
3	Analyst:	HM						
4	Prep Chemist:	LW		**MDL = 0.47 ng/g**				
5	Data Entry:	WA		**RL = 3.33 ng/g**				
6	QC/Data Checks:	SM		**All data is based on wet weight.**				
7								
8				**Hg Conc. (ng/g)**				
9	**Analysis Date**	**Batch #**	**Sample ID**	**CAS#7439-97-6**	**Qualifier**	**% Moisture**	**Qualifier**	
10	9/15/2009	HG-005	521729	296.99		78.73		
11			529039	156.84		88.81		
12			529119	217.82		82.71		
13			529079	21.41		76.88		
14			526419	170.16		80.58		
15			529209	121.15		75.47		
16			527209	853.72		79.49		
17			530399	283.96		80.20		
18			526319	86.45		93.08		
19			532589	449.61		80.91		
20			532879	493.04		81.21		
21			525119	175.39		96.07		

Final Data | QC | Calibration | Sequence | Moisture Analysis

Reporting Example (cont.) – PPCP Data Format



	A	B	C	D	E	F	G	H	I	J	
1	CAS Number	Name	MDL	QL	blank-extraction	Q	Site 009	Q	Site 010	Q	Site
2	51481-61-9	cimitedine	0.6	1.9	nd		nd		nd		
3	66357-59-3	ranitidine	3.5	11	nd		nd		nd		
4	738-70-5	trimethoprim	0.8	2.5	nd		nd		nd		
5	723-46-6	sulfamethoxazole	0.5	1.6	nd		31.1		34.9		
6	64520-05-4	10-hydroxy-amitriptyline	0.2	0.6	nd		nd		nd		
7	58-33-3	promethazine	0.4	1.3	nd		nd		nd		
8	110429-35-1	paroxetine	0.5	1.6	nd		nd		nd		
9	28981-97-7	alprazolam	2.9	9.1	nd		nd		nd		
10	549-18-8	amitriptyline	0.2	0.6	nd		nd		nd		
11	86-13-5	benztropine	0.5	1.6	nd		nd		nd		
12	83891-03-6	norfluoxetine	2.3	7.2	nd		nd		nd		
13	59333-67-4	fluoxetine	0.9	2.8	nd		nd		nd		
14	79902-63-9	desmethylsertraline	3.0	9.4	nd		nd		nd		
15	79559-97-0	sertraline	0.9	2.8	nd		nd		nd		
16	18559-94-9	albuterol	3.1	9.7	nd		nd		nd		
17	29122-68-7	atenolol	1.9	6	nd		1.9	E5	4.1	E5	
18	4205-91-8	clonidine	11	35	nd		nd		nd		
19	124-90-3	oxycodone	0.8	2.5	nd		nd		nd		
20	51-63-8	amphetamine	0.5	1.6	nd		nd		nd		
21	143-71-5	hydrocodone	1.2	3.8	nd		nd		nd		
22	396-01-0	triamterene	0.4	1.3	nd		nd		nd		
23	56392-17-7	metoprolol	4.3	14	nd		8.7	E5	6.3	E5	
24	76095-16-4	enalipril	0.3	0.9	nd		nd		nd		
25	318-98-9	propranolol	1.4	4.4	nd		nd		nd		

analytical batch summary | antis | cardios | neutrals | acidics | sequ

Part 1: Quiz #1

Why do you have to implement QA?

a) It ensures our decisions are transparent

b) It ensures our decisions are scientifically and legally defensible

c) It ensures that taxpayer money is well spent

d) It is EPA policy

e) Because my QA Officer said so

f) All of the above

Part 1: Quiz #2

True or False?

1. The Data Quality Act requires EPA to ensure the quality of all information it disseminates
2. If EPA gives money to another organization (e.g., a state or university), it can avoid the DQA
3. If you do systematic planning and write a QAPP, then you have complied with the Act
4. If you use data that someone else has already published, then you can assume that the data are of acceptable quality

Agenda- Part 2

Part 2: Data Review (verification/validation)

- Scope
- Philosophy
- Details
  - Existing Data
  - Batch and Batch QC data
  - Approaches
- Flags
- Examples

Estimated Time for Part 2:
50 minutes
(followed by a short break)

Data Review - Scope

Data review process is applied to all data types

- New and existing data
- Field and lab data
- Customize per method, study, and regulatory requirements

Data Review - Philosophy

- Some organizations believe that the goal of data review is to find fault with and eliminate as much data as possible
- *Better philosophy:* Maximize data usability by understanding the effect of the shortcomings on the results and on the environmental decision to be made
- *Best philosophy:* Prevent/minimize data quality problems through careful planning, and maximize data usability by understanding the effect of the shortcomings on the results and on the environmental decision to be made

Using the "Maximize Data" Philosophy

- Avoids delays for resampling and reanalysis

- Saves money

- Difficulty is in deciding what data can be salvaged and determining how results are affected.

  - Experts are those with data gathering and review experience
  - Knowledge of problems specific to an analysis are with analyst(s) who collected the data – but be wary of attempted snow job
  - Document findings (transparency to data user & public)

- Final decision on data usability belongs to decision maker or data user

Existing Data – Data Review/ Data Quality Questions

- What are the minimum requirements for use of these data?
- Were the data generated in a way that meets the quality criteria for the project?
- Do the data have metadata describing the data and quality criteria?
  - If yes, do they meet method criteria and/or your project-specific criteria?

EPA United States Environmental Protection Agency

Existing Data - Data Review/ Data Quality Questions (cont.)

- Are the data relevant, representative, and comparable?
  - Target populations
  - Methods
  - Measurement errors
  - Locations
    - Spatial or depth differences
    - Spatial or grab composites
  - Timing
    - Sampling seasons
    - Time of day
    - Grabs vs. temporal composites
  - Detection, quantitation & reporting limits

Confidence in Existing Data

Consideration	High Confidence	Low Confidence
Level of peer review	High level of review	Limited peer review
Accessibility	Widely available to the public	Difficult to obtain (e.g., draft reports, unpublished data)
Reproducibility	Results can be reproduced or methodology can be followed by others	Results cannot be reproduced, or methodology hard to follow
Focus	Focus on factor of interest	Characterize a related factor
Source	Direct observation or measurement	Not direct measurements

Confidence in Existing Data (slide 2)

Consideration	High Confidence	Low Confidence
Pertinence	Focused on the system of interest	Not relevant to system of interest
Project planning	Generated under a document quality system	No evidence of a quality system
Data quality assessments	Data were assessed against stated DQOs	Data were not assessed
Validity of approach	Used best available methodology	Serious limitations to the approach
Bias in study design	Potential biases are stated or can be determine from design	Study design introduces bias into the results

Confidence in Existing Data (slide 3)

Consideration	High Confidence	Low Confidence
Number of studies or sources	3 or more	Only 1
Agreement between sources	Different sources agree	Different sources do not agree
Age of data	Represent current conditions or practices	Older data, not representative of current conditions or practices
Temporal representativeness	Data collection period provides representativeness	Data were not collected over sufficient time frame
Study size	Large volume of data generated	Limited data generated

Data Review Process

- Often varies by organization, study, individual
  - Recommend standardizing process within a study
  - Minimizes variability over time
  - Minimizes variability between staff
  - Tools: SOPs, Guidelines, Checklists, Automated procedures
- Starting early enhances ability to maximize data
  - Increases odds of being able to reanalyze if needed
  - Reanalysis is cheaper than re-sampling

Data Review Process (cont.)

- Standardized, multi-step process facilitates comprehensive, timely review
- Data Completeness Check
  - Confirms all requested analyses were performed
  - Confirms all required deliverables were submitted (hardcopy format, electronic format, required elements and units
  - Doing this first avoids wasting time with in-depth review of incomplete data
- Instrument Performance Check
  - Verifies that instrument is properly calibrated and contaminant free
  - Were initial calibration, calibration verification checks and calibration blanks performed at correct frequency and do they meet your performance criteria

Data Review Process (cont.)

- Lab Performance Check
  - Verifies that lab performed the analytical procedures correctly with acceptable precision and accuracy
  - Examines
    - Holding times (sampling, extraction/digestion, analysis)
    - Initial and ongoing precision and accuracy tests
    - Preparation blanks (aka “Method Blanks”)
    - QC samples (e.g., LCS, LFB, OPR, SRMs)
    - Media Sterility Checks
    - Positive/Negative Control results
    - Incubation length & Incubation temperature
  - Should address both frequency and performance criteria

Data Review Process (cont.)

- Method/Matrix Performance Check
  - Helps discern whether QC failures are associated with lab or method performance vs. matrix complexities
  - Examines
    - Results of spiked field samples
      - Matrix spikes, matrix spike duplicates, surrogate spikes, isotopically labeled compounds
      - Percent recovery (accuracy) and relative percent difference (precision)
    - Clean-up procedures and dilution levels
      - Were appropriate dilutions and sample cleanups performed?

Data Review Details – Completeness

- Is the data package complete? (hopefully you defined required elements during planning!)
  - Results for all samples, including all QC samples (completeness)
  - Ideally includes
    - Traffic reports or chain-of-custody forms
    - Narrative
    - Summary result forms
    - Raw data
    - Logs and bench notes
  - Electronic data deliverable (EDD) if required

Data Review Details – Sample and Batch QC data

- Individual samples
  - Results for target analytes
  - Surrogate recoveries (labeled compound recoveries for isotope dilution)
    - Demonstrates that method and lab perform the same way on samples as with reference matrices
- Supporting Batch QC Data
  - Prep batch
    - Demonstrates the preparation processes were in control
  - Analysis batch
    - Demonstrates the analysis processes were in control
  - Not always the same 

Data Review Details – Sample data

- Results for target analytes
  - Verify holding times were met (sample, extraction, analysis)
  - Verify reported results were within calibration range
  - Verify appropriate dilutions were performed
- Surrogate or labeled compound recoveries
  - Verify it was performed in every sample and QC sample
  - Verify recoveries meet pre-defined acceptance criteria
- Method-specific details
  - Ion abundance ratios, signal to noise, confirmation analyses, etc.

Data Review Details – Prep vs. Analysis Batch QC Data

- Prep Batch
  - Group of similar samples carried through all sample processing steps together using the same techniques
    - Usually defined as a group of 10 or 20 samples prepared on same day or shift
    - Includes digestion/extraction/clean-up steps
- Analysis Batch
  - Group of samples, extracts, or digestates (including QC samples) analyzed together on the same instrument on the same day or shift
    - Usually defined as a group of 10 or 20 samples analyzed on the same day or shift
    - Some analytes (e.g., VOCs) may only have an analysis batch

Data Review Details – Prep Batch QC

- Prep Blank (aka Method Blank)
  - Demonstrates contamination is not introduced during the sample preparation or analysis processes
  - Verify frequency and absence of “hits”
    - Ideally non-detect
    - If detected, should be small fraction of associated sample results
- Lab control sample (LCS)
  - aka Lab Fortified Blank (LFB) or Ongoing Precision and Recovery Sample (OPR)
  - Demonstrates that laboratory is in control
  - Verify frequency and ability to achieve pre-defined acceptance criteria

Data Review Details – Prep Batch QC (*cont.*)

- Matrix Spike and Matrix Spike Duplicate
  - Demonstrate that method is applicable to the particular matrix
  - Usually applied to a representative sample matrix in the same prep batch
    - Typical frequency = 5 or 10% of samples in batch (e.g., 1 per 20 or 1 per 10)
  - Must be applied to every different matrix, not the best looking sample out of samples received
  - Examine frequency, recoveries, and RPD and evaluate against pre-defined performance criteria

Data Review Details – Analysis Batch QC data

- Instrument and Equipment Blank(s)
  - Demonstrates freedom from contamination
  - Calibration blanks
- Instrument Calibration
  - Initial calibration
    - Verify it was performed as required by method (e.g., 3-point, 5-point, etc.)
    - Verify it was performed at required frequency (e.g., every 8 hours, daily, or only when needed)
    - Verify sequence (e.g., after tune, before calibration blank, etc.)
    - Verify linearity requirements were met
  - Calibration Verification (or continuing calibration)
    - Verify frequency, sequence, and performance criteria were met

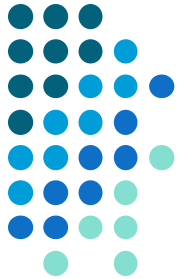
Data Review Details – Analysis Batch QC data (*cont.*)

- May be method-specific, e.g.,
  - ICP Metals: Interference Check Samples, Inter-element Correction Factors, Serial Dilutions
  - Low level mercury: Bottle blanks
  - HRGC/HRMS Dioxins: Window-defining Mixture, Isomer Specificity Standard
  - Volatile & semivolatile organics: instrument tune (e.g., BFB, DFTPP)
- Method-specific details are beyond scope of this course
  - Must be familiar with method

Data Review – Detailed (Batch QC data cautions)

- All samples in a prep batch do not have to be analyzed in the same analysis batch
  - Identifying the QC "failures" is easy (e.g., list out or flag all QC samples that failed to meet specified criteria)
  - Linking QC failures to their associated samples is hard
    - Especially if the prep batches and analysis batches don't match
    - Many programs imply they link but they don't (or only link some things)
- Linking QC failures to samples supports data usability
- Flagging QC failures (not linking to samples) supports data transparency

Linking Batch QC to Affected Samples



- If problem with LCS/LFB/OPR; e.g., low recoveries, loss likely in prep/extraction/cleanup. Samples should be re-prepped.
- If problem is elevated blank value and result in associated sample is < 10x blank value (ideally, <20x), difficult to distinguish sample result from possible contamination. If critical, samples should be re-prepped after contamination eliminated.
- If problem with MS/MSD*; e.g., low recoveries, and LCS/LFB/OPR is OK, problem is likely with sample matrix. If critical, samples should be re-analyzed, or consideration should be given to recovery correction
  - * Important to know if MS/MSD is representative of all samples in the batch

Linking Batch QC to Affected Samples (example)

- Data package has results for 12 samples
  - All data were in the same instrument run
- Prep Blank shows signs of contamination. QAPP says:
  1. Flag all samples associated with high prep blank
  2. Determine if the high blank impacts the samples
- Do you flag all 12 samples?
  - Not necessarily!
  - You need to figure out if the 12 sample were prepared in the same batch as the prep blank
  - If you have other samples in your project (from a different analysis run), you also have to figure out if they were prepped in same batch as this blank

Field QC

- Important not to overlook field QC
  - Often handled during data quality assessment stage
  - Requires up-front planning of who examines what
- Example field QC elements
  - Field duplicates, splits, replicates (frequency, recovery, and precision)
  - Blanks
    - Trip blanks – usually associated with volatile analytes of interest e.g., VOCs, mercury
    - Field blanks
    - Examine frequency and closeness of “hits” to associated sample results
  - Sample holding times (from collection to analysis), which may be beyond the lab’s control

Data Integrity

- Depends on integrity of project, field, laboratory, and management people
  - Sloppiness
    - Mislabeling of sample containers
    - Failure to conduct all analytical steps (e.g., cleanup)
  - Improper field practices
    - Altering sampling location to hide possible contamination problem
    - Refilling of VOA sample bottle at a later time when bubble discovered
    - Collecting sample from a location accessed easily rather than from designated location
    - Filling sample bottle with reagent water to avoid violation
- * Note that “Integrity” in this context is different than in the Data Quality Act context

Data Integrity (*cont'd*)

- Improper laboratory practices
  - Altering receipt or analysis date
  - Failure to analyze samples (dry labbing)
  - Manipulation of sample prior to analysis (e.g., juicing)
  - Double injection to increase recovery of MS/MSD
  - Manipulation of BFB or DFTPP spectral intensities to meet criteria
  - Including background area to meet calibration verification criteria
  - Post-analysis changing of data to meet contract specifications
- Items a reviewer/validator can look for:

  - Inconsistent dates in data package (e.g., was the sample analyzed before it was extracted?)
  - Manual integration in QC data files to meet specifications

http://www.epa.gov/quality/qs-docs/g8-final.pdf

Data Review Tools

- Checklists
- Guidelines
- SOPs

  - Should be tailored to method, program, and/or project
  - Examples provided in handouts

- CLP Data Assessment Tools (next slide)

CLP Data Assessment Tools (DAT)

- Data Verification via
  - Automated Lab Self-inspection
  - Automated Contract Compliance Screening (CCS)
    - CCS output comes in the form of a “defect” statement
- Data Validation via
  - Metals: Computer-Aided Data Review & Evaluation (CADRE)
  - Organics: Electronic Data Exchange and Evaluation System (EXES)
  - Output is a flag and a description of the problem
- In practice, there is much overlap (true of many programs)
  - Contracts requirements reflect basic data usability needs

What CLP DAT Does/Does Not Do

- Rapid, standardized data review process
- Covers verification/validation only
  - Does not recognize site- or project-specific modifications or quality objectives
  - Doesn't examine chromatograms and mass spectra
  - Is not data quality assessment
- Generally, one size fits all
  - National Functional Guidelines (NFGs) are highly method-specific and not site-specific
  - NFGs implemented through automated routines
    - Computers can't apply best practical judgment (BPJ)
    - Needs to be supplemented with manual validation

A Word of Caution

In any program:

- Contractually compliant data may not be usable
- Contractually non-compliant data may be usable
- Hence, you have to *assess* the quality of your data against your project needs
  - *See Part 3, Data Quality Assessment*

Data Flags - Overview

- Flags (data qualifiers) are typically used to indicate something unusual or incorrect with a result
  - Can be descriptive (e.g., “U” means undetected)
  - Can reflect data quality issues (e.g., “B” means the analyte was detected in the associated method blank)
  - Can be interpretive (e.g., “RH” means the result may have a high bias due to a QC failure)
- Some flags require evaluation/interpretation by someone with specialized knowledge (e.g., the analyst, a senior reviewer with strong subject-matter experience)
  - Example: In the CLP semivolatile organics SOW, flag “A” is for a compound suspected of being an aldol-condensation product. If we really want to know what the compound is, we must consult with experts.

Data Flags – Meaning

- Many flags in some programs
  - CLP organics and inorganics use “U”, “J”, “N”, “P”, “C”, “B”, “E”, “D”, “S”, “X”, “Y”, “Z”, and “+”
- Flags can have different meanings in different programs
- Flags can have different meanings in the same program
  - “N” in CLP semivolatile organics analysis means presumptive definition of a tentatively identified compound
  - “N” in CLP inorganics (metals) analysis means the spike recovery is not within 75 – 125 %
- Flags are applied at different levels
  - Lab-applied flags
  - Data reviewer-applied flags

Data Flags – Impact on Decisions

- Common flags of concern in environmental decisions
  - "U" (not detected or not quantified)
    - If the detection limit is above the action level, either the wrong method was chosen or interferences precluded detection
    - Choosing a method sensitive enough is part of the DQO/MQO/QAPP process; overcoming an interference is a lab/project responsibility
  - "J" (estimated): requires an estimate of uncertainty in order to support an environmental decision

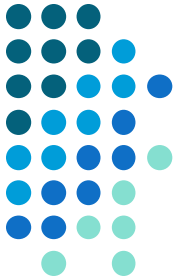
Data Flags – Impact on Decisions #2

- Common flags of concern (cont'd)
  - "B" (compound found in blank): Indicates that the result for that compound in an accompanying sample could be a false positive (rule of 5x or 10x)
  - "E" (concentration exceeds calibration range): affects results only if accurate concentration must be known (not just exceedance of a level)
  - "X", "Y", and "Z": mystery flags specific to a laboratory, an analysis, or data

Overarching Data Review Issues

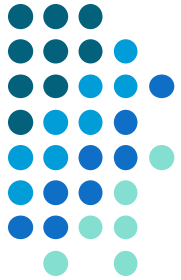
- Many reviewers accept data without question
  - No “reality check”
  - No attempt to reconcile with project objectives
  - No concern for “does it make sense?”
  - No attempt to reconcile with other data (primary or secondary)
  - No checking with experts
- Many reviews don’t include all error sources
  - CLP reviews do not include field data
  - Sample location may be incorrect
  - Sampler may have estimated or could estimate sampling error
  - Flow meter may be out of calibration
  - Site hydrology may not have been considered

Example: EPA Detection/Quantitation Limit Study



- Highly visible study part of a multi-year effort
  - Follow-on to a Federal Advisory Committee effort
  - Potential to impact future regulations
  - Focus is on testing method detection limit and quantitation limit procedures (not on characterizing pollutants)
- Data review needed to verify
  - Analytical requirements were met (e.g., methods properly followed and required QC elements were performed)
  - Required procedures for determining detection and quantitation limits were followed
  - Calculations were properly performed
- Overall Strategy
  - Build on lessons learned from an earlier (larger) study
  - Automate where possible, but manually review the 'hard to automate' aspects (e.g., analytical sequence, initial calibration)

Example: EPA DQ Limit Study – Data Reporting Strategy



- Calculations performed to determine limits
  - Simple spreadsheet with specified data elements but format of lab's choice
- Study sample and supporting QC data results
  - Excel file with 24 specific fields and specific codes for populating each field (e.g., standard QC element names, standard analyte names)
  - Hardcopy submission of data narrative, run chronology, initial calibration data and raw supporting data

Example: EPA DQ Limit Study – Data Review Strategy

- Use:
  - Standardized data review checklists (Word) tailored to reflect both method and study requirements
  - Experienced chemists to review the chemistry components of the data
  - Experienced statistician to verify that the det/quant methodology was properly followed and calculations properly performed

Adobe Acrobat Document

Example: PPCP Data

- Evaluate Pharmaceuticals and Personal-Care Products (PPCPs) in wastewaters to support possible effluent regulations
- EPA Method 1694: PPCPs by isotope dilution liquid chromatography/tandem mass spectrometry (LCMSMS)
  - 4 analysis groups – based on characteristics of PPCPs
    - 3 groups by positive electrospray ionization (ESI+)
    - 1 group by ESI-
- Method still in development phase at time of study

PPCP Data Reporting and Data Review Approach

- Use standard EPA EAD data reporting format
  - EDD plus hardcopy reporting of ICAL and raw/instrument data
- Use Excel-based, standardized data review checklists
- Why?
  - EAD EDD flexible enough to handle new method/analytes
  - Electronic reporting of ICAL and raw data is hard to set up
  - Excel-based checklists:
    - Flexible enough to be tailored to meet project-specific needs
    - Document the reviewer findings and allow reviewers to partially automated review processes
  - Cheaper than fully manual review
  - Cheaper than building fully automated review routines

PPCP Data for Each Sample

Received in Excel spreadsheets per EDD specs. Results for each analyte in each sample included:

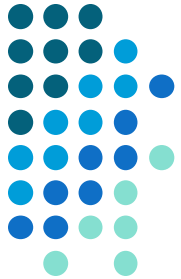
- Sample Episode number
- EPA sample number
- Lab sample number
- Lab name
- Lab contract number
- Data entry date and time
- Sample matrix
- QC code
- Sample date
- Date received at lab
- Analysis performed (method)
- Sample size and unit
- Extraction date
- Analysis date and time
- Extract volume
- Injection volume
- Instrument identifier
- Calibration date

PPCP Data for Each Sample (cont.)

- LC column identifier
- LCMSMS data file identifier
- Calibration verification file identifier
- Method blank identifier
- Batch identifier
- Laboratory's name for compound
- EPA's name for compound
- CAS Registry Number
- Concentration found and unit
- Dilution factor
- For labeled compounds
  - Upper and lower QC acceptance criteria
  - Relative retention time
  - Ion abundance ratio
- For matrix spike and duplicate
  - Spike added
  - Spike measured
  - Spike recovery

Calibration and raw data were reported in hardcopy

PPCP Data Verification/Validation Process



- Reviewer used an Excel-based checklist to evaluate the data
  - "Standard" checklist format, but customized to reflect each method
  - If project specs require different MQOs than specified in method, checklist is modified to reflect project specs
- Use of Excel allowed the reviewer to:
  - Import data from the EDD into a "data review" file that contained QC acceptance criteria
  - Sort and filter data as needed to review each element (e.g., examine only the Cal Ver data for Acetaminophen)
  - Quickly identify data that were outside the criteria
  - Maintain records of data review in the same file as the completed data review checklist sent to the client (different worksheets)

PPCP Data Package – Checklist

- Checklist evaluates the following elements (as yes/no decisions):
  - Traffic Reports complete
  - Samples received at < 6 °C
  - MDL study complete
  - IPR study complete
  - Hardcopy data package complete
    - Narrative
    - Summary results forms
    - Raw data
    - Logbook data
  - Electronic data deliverable in form specified
  - Results received for all samples and QC samples and match Traffic Report
  - Result present for each contract-specified analyte

PPCP Data Package Checklist (slide 2)

- Aqueous samples extracted within 7 days of collection and analyzed within 40 days of extraction
- IPR performed and within criteria
- Calibration performed and linearity within criteria
- Each analyte quantified against correct reference
- Signal-to-noise ratio (S/N) is greater than 10 for each analyte in calibration verification
- Retention time (RT) for latest eluted analyte in each group is greater than RT specified
- Calibration verification(s), OPR(s) and blank(s) performed at required frequency, prior to analysis of samples, and within criteria
- Analytes not detected in blanks
- Labeled compounds within QC acceptance criteria in samples, OPRs, and blanks

PPCP Data Package Checklist (slide 3)

- Reviewer comment column - cell for each item reviewed allows reviewer to state the impact of specifications not met, or on data usability. Examples:
  - Laboratory used weighted linear regression and correlation coefficient of >0.985 rather than <20% RSD of relative responses (isotope dilution) or <35% RSD of response factors (internal standard), as specified in method. Some analytes did not meet >0.985 spec.
  - Disposition: The error in results for the affected analytes will usually be increased by the increased error in calibration. This increased error should be allowed for in any environmental decision. An estimate of the amount of the error increase, by analyte, can be provided, if desired.

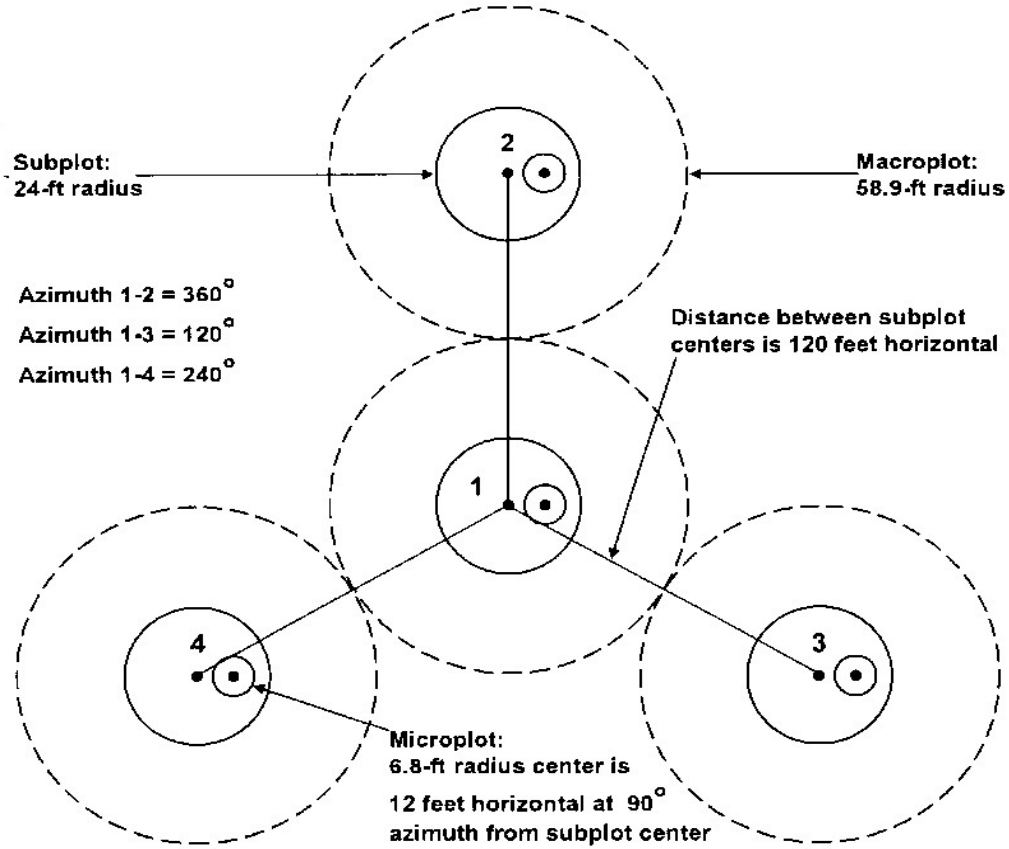
U.S. Forest Service (USFS) – Forest Inventory and Analysis (FIA)

- Provides information needed to assess environmental quality of Nation's forests
- Measurements made by surveying number and characteristics of trees in check plots; e.g.:
  - Species
  - Height
  - Diameter
- High-quality studies with high-quality data

Technical Report RMRS-GTR-181

Resource Bulletin NRS-41

USFS FIA Measurement Plot



EPA United States Environmental Protection Agency

USFS FIA (cont.)

- Quality assessed by comparison of survey results by two teams
  - Teams totally blind to each other
    - Different days, personnel, equipment
  - MQOs established as:
    - Percent agreement (usually 90%, 95%, 99%)
    - Tolerance on agreement; e.g., +/- 10%
  - Data compared to see if MQOs are met
    - Discrepancies noted
    - Surveying improved as result

Example Data Issues from Great Lakes Legacy Act Projects

- Project “Alpha”: Total PCB results reported were low by a factor of 10
  - Identified through recalculation of Total PCB results
- Project “Bravo”: Surface and subsurface sample results switched for one station. Sample IDs non-unique
  - Identified by evaluating ranges of results by depth and noting surface samples were typically lower concentrations
- Project “Charlie”: Incorrect location information (e.g., samples all off by 50 feet due to conversion error)
  - Identified through plotting of sample locations
- Multi-project issues: Missing work orders, QC batch information, reporting limits, and other technical details

Example – Project "Delta" (slide 1 of 2)

- Sediment Remediation Project that required quick turnaround analyses and interpretation during remedial activities
- Mercury levels in post-dredge samples so high that lab experienced extensive contamination of facility and equipment
  - Slowed down lab turnaround
    - Emphasis in QAPP and better communication would have alerted lab

Example – Project "Delta" (*cont.*)

- Laboratory did not review QAPP until after analyses were performed
  - Lab was not aware MS/MSDs were randomly selected in sampling design and did not complete required frequency of MS/MSDs
  - After review, lab commented that QC acceptance criteria in the QAPP were incorrect
    - When it was noted to the lab that the QAPP criteria were from the lab's SOP, lab responded that the SOP was outdated
    - Lab's SOPs were the basis of the data review, which in turn, had to be readjusted

Project “Echo” Data Issues

- Sediment remediation process with several contaminants of concern, including benzo(a)pyrene (BAP)
  - EPA Method 8270 was selected for BAP analysis during the project planning
  - Lab analyzed for BAP using Method 8270 as per the lab’s SOP
- During data review, it was noted that the lab used pyrene instead of BAP as the spiking compound in their matrix spike and lab control samples
  - BAP and pyrene do not behave similarly
  - Lab should have spiked with BAP instead of pyrene
- What happened?
  - Review of lab SOP indicated that the lab was using an older version of Method 8270
- Lessons Learned
  - Specify the exact version and dates of the procedures you plan to use during planning and in the QAPP
  - Prepare/review SOPs for data gathering activities before data gathering begins

PCB Data for Delaware River TMDL

- Delaware River Basin Commission (DRBC) charged with developing a TMDL for PCBs in the Delaware River
  - Fish advisories showed Delaware impaired by PCBs
  - Remedy is to develop a total maximum daily load (TMDL)
  - Ambient criteria range from 8 to 45 pg/L, depending on location
  - EPA Method 1668A selected as method best capable of measuring to these levels
  - Concentrations measured ranged from approx 200 pg/L to 6000 pg/L, depending on location
  - DRBC developed its own data-review guidance.
    - Supplements EPA OW guidance and QC in Method 1668A
    - Addresses larger sample volume (2 liters)
  - Slides that follow highlight modifications made for project

    http://www.state.nj.us/drbc/PCB_info.htm

DRBC PCB TMDL

- Project-specific definitions and flags
  - Data qualifier flags
    - Flags agree largely with CLP
    - Additional flags
      - X = Result is from re-extraction and/or reanalysis
      - EMPC = Estimated maximum possible concentration – Congener was detected but did not meet identification criteria

http://www.state.nj.us/drbc/PCB-DataQualFlags.pdf

  - Reporting rules for co-eluting congeners
    - Cxx = Congener co-elutes – Result is reported under the lowest numbered PCB congener in the co-elution, where xx is the number of the lowest co-eluted congener
    - Result will be identified with a C in the flags column

http://www.state.nj.us/drbc/PCB-CoelutingCongeners.pdf

DRBC PCB TMDL (slide 2)

- Project-specific definitions and flags (cont'd)
  - Method blank contamination decision rules
    - Decision tree based on level of congener in blank and associated sample, and total PCB concentration in sample

    http://www.state.nj.us/drbc/PCB-MethodBlankRules.pdf
  - Rinsate blank decision rules
    - Based on contamination being < 40 pg/L for any congener
    - Total PCB contamination cannot exceed 600 pg/L

    http://www.state.nj.us/drbc/PCB-Rinsate.pdf

    Estimated detection limit (EDL) decision rules
    - Based on formula using S/N in region of chromatogram where peak should appear

    http://www.state.nj.us/drbc/PCB-EDL.pdf

DRBC PCB TMDL (slide 3)

- Examples of project-specific modifications
  - Collect 2-L samples in duplicate
    - Backup for re-extraction if necessary
  - Laboratory must supply reagent water for trip blanks
    - Ensure accountability
  - Do not filter sample – extract total 2-L volume
    - Avoid complexity (solids are expected to be <1%)
  - Minimum retention time spec for PCB 209 may be waived if justified
    - Allows for newer GC technology; e.g., electronic pressure control
  - All extracts in previous shift must be reanalyzed if MS resolving power check is not met at end of shift
    - Assures adequate mass resolution for all samples
  - Ion abundance and signal-to-noise ratio (S/N) specs must be met with 0.5 ng/mL standard
    - Verifies sensitivity

DRBC PCB TMDL (slide 4)

- Reinforcement of QC specs in Method 1668A
  - Calibration must meet linearity spec for isotope dilution (<20% RSD) and internal standard (<35% RSD).
    - Preclude wild calibration point(s)
  - Calibration verification
    - Verification result for each compound must be within QC acceptance criteria in Method
      - Assures instrument remains in calibration
    - Each compound must have S/N > 10
      - Assures adequate sensitivity to PCBs
    - Retention times (RTs) for labeled compounds must be within ±15 seconds of RT in calibration or verification
      - Assures RTs haven't shifted
    - Relative retention times (RRTs) must be within limits in Method
      - Assures proper congener identification
  - Blank *(click link)*

Example: Reviewing Volunteer Data

- RiverWatch (WR) Program
  - Led by Friends of the Fox River
  - Volunteer sampling activities – are the data usable?
- Information submitted
  - Samples collected but no data sheets submitted
  - Samples collected but not properly preserved – specimens degraded
  - Volunteers who weren't listed as having been trained
  - Samples and data with verifiable information
    - 99 verifiable units
    - Half were re-evaluated by expert (Data QA)

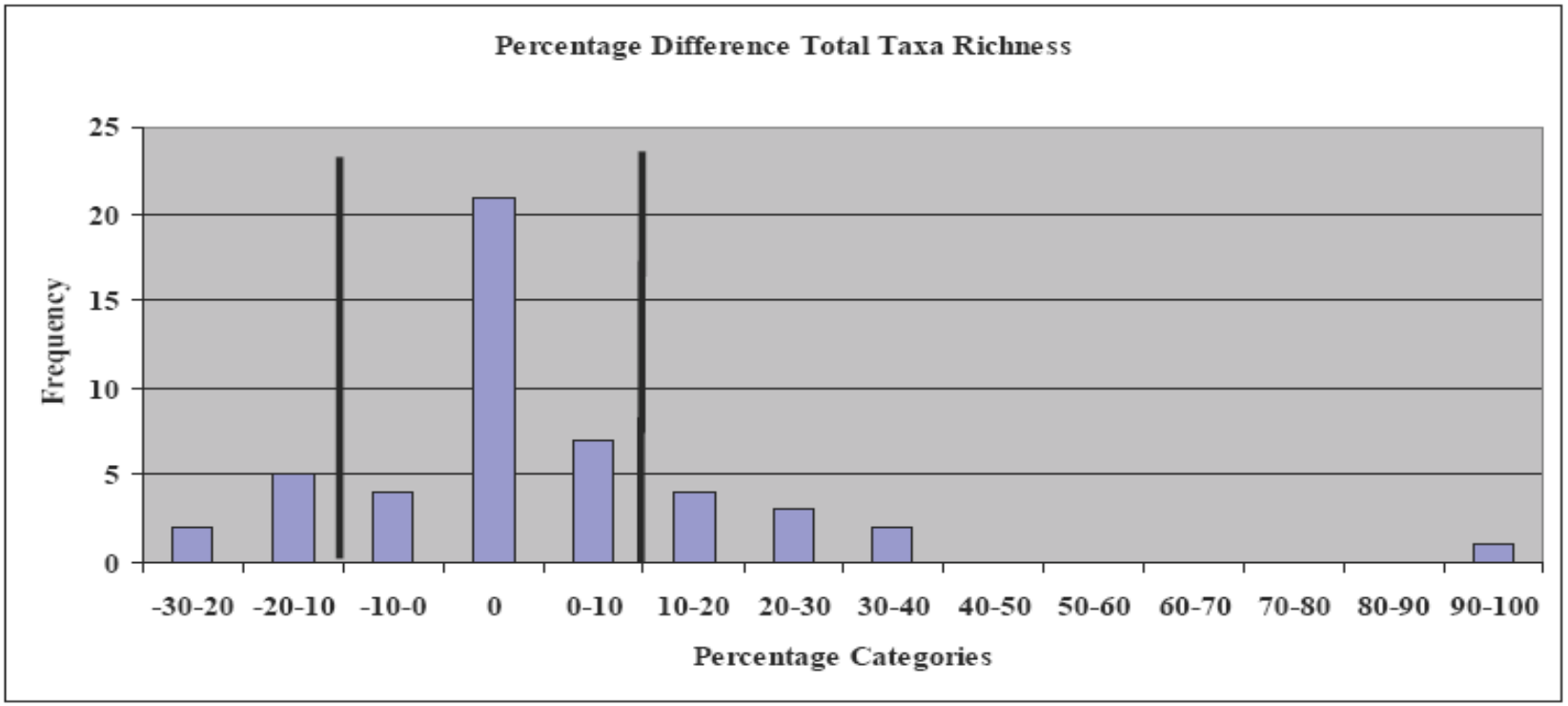
Could not be verified; data were eliminated

Example: Reviewing Volunteer Data (continued)

- Data Collected
  - Taxa richness and Macroinvertebrate Biotic Index (MBI) values
- Data Evaluation Techniques
  - Re-examination of samples by expert
  - Comparison of expert findings to volunteer findings
    - Averages
    - Ranges
    - Frequency histograms
- Data Evaluation Findings
  - Averages and ranges often implied agreement between volunteers and experts
  - Histograms yielded a more nuanced picture

Example: Reviewing Volunteer Data (continued)

Total Taxa Richness Data Reported by Volunteer and QA Expert		
	Average (taxa)	**Range (taxa)**
Volunteer	10.0	2 – 19
Expert	9.7	2 – 16



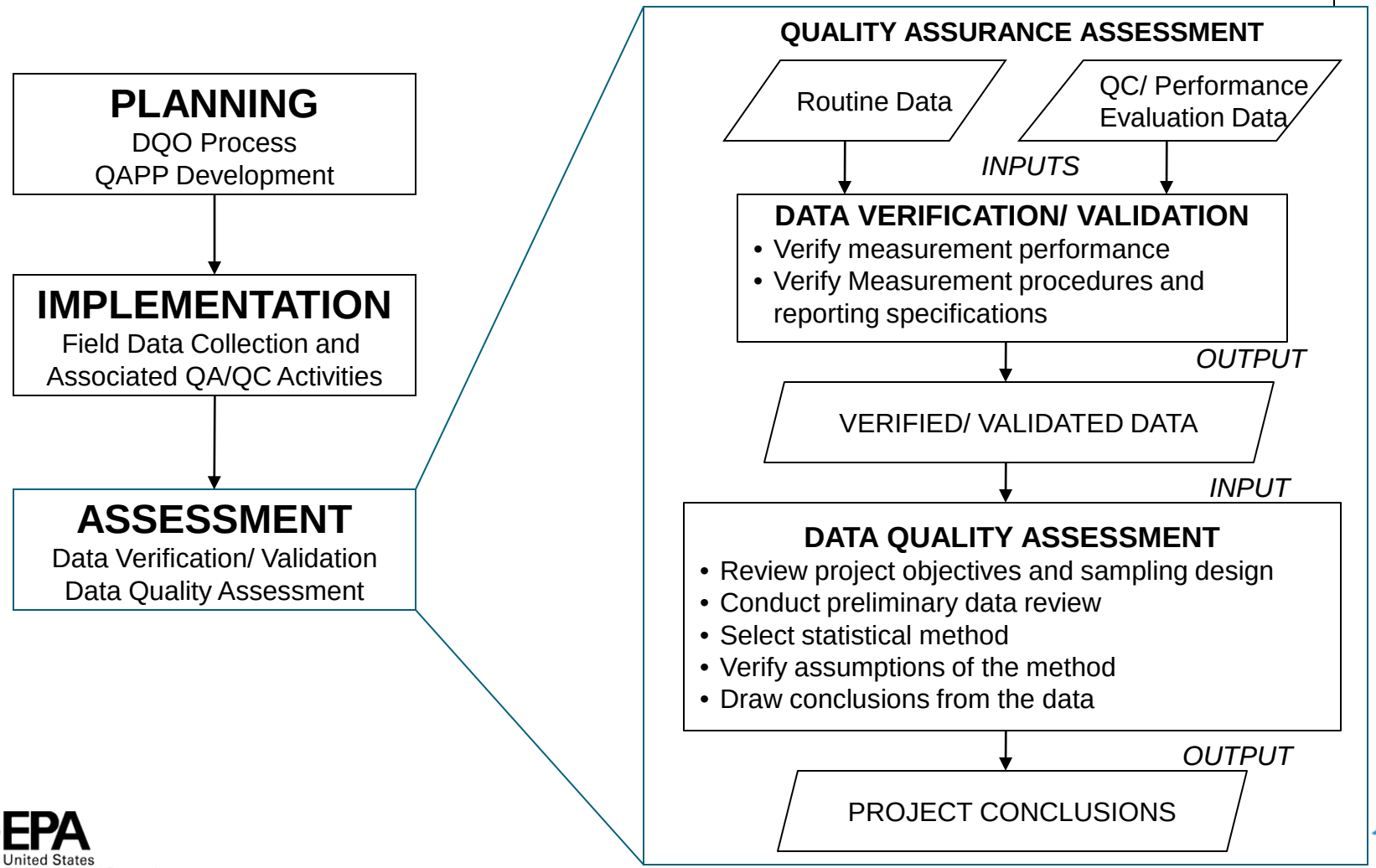
Agenda- Part 3

Part 3: Data Quality Assessment

- Role in Data Life Cycle
- Refresher– distinguishing between
  - Data Verification
  - Data Validation
  - Data Quality Assessment
- Data Quality Assessment Tools
- General Assessment Factors
- Five Steps of Data Quality Assessment
- Data Suitability

Estimated Time for Part 3: 30 minutes

Data Life Cycle



EPA United States Environmental Protection Agency

Data Verification - Summary

The process of evaluating the completeness, correctness, and conformance/compliance of a specific data set against the method, procedural, or contractual requirements

- Goal: Ensure and document that the data are what they purport to be (reported results reflect what was done)
- Is it complete/did they do what they were supposed to?
  - All required samples and QC were collected and analyzed
  - Data were reported in the correct format
  - All required data elements are present
  - All QC samples met pre-defined performance criteria or failures were clearly identified
- Can you recreate the results from the raw data?
  - Reproducibility required for “influential” data under DQA

Data Validation - Summary

An analyte- and sample-specific process that extends data evaluation beyond method, procedural, or contractual compliance (i.e., data verification) to determine the quality of a specific data set

- Goal: Evaluate whether data quality goals were achieved for the data set
- Examines failures and potential impacts in field and lab for individual sampling events or data "packages"
  - Precision
  - Representativeness
  - Comparability
  - Accuracy
  - Completeness

Data Quality Assessment - Summary

A statistical and scientific evaluation of the data set to determine the validity and performance of the data collection design and statistical test, and to determine the adequacy of the data set for its intended use

- Goal: Focus on environmental decision making, asking if the data sets generated can effectively and credibly support the decisions
  - Data verification and validation focus on specific sampling and analysis processes and results (not decisions)
- Assesses overall
  - Precision, accuracy, representativeness, completeness and comparability of data
  - Achievement of data quality goals for the project

Data Quality Assessment – Questions to be Answered

- Can the decision (or estimate) be made with the desired level of certainty, given the quality of the data?
  - Addresses user's immediate needs
- How well did the sampling, analysis, review, and validation systems perform against requirements for the project and in the QAPP?
  - Addresses robustness of the pertinent systems
- Is is likely that sufficient samples were collected and analyzed to enable a user to see an effect if it was present?
  - Addresses that sufficient data were gathered to support the decision
- If the sampling, analysis, review, verification, and validation systems are used again, can the data be expected to support the same intended use with the same level of certainty?
  - Addresses whether this project is unique

Example #1 – Distinguishing among the Terms

After a contaminated sediment remediation project, the site is re-sampled for PCBs (the contaminant of concern)

- **Data verification** documents that PCB recoveries in a spiked sample were below control limits
- **Data validation** determines
  - The cause for non-conformance was probably a low spike amount relative to the background sample concentration
  - All other matrix spike and LCS recovery criteria were met and field duplicate RPDs were acceptable
- **Data quality assessment** considers that
  - All post-remediation sediment samples at the site had PCB concentrations well below the action limit for the site
  - All aspects of the study design were implemented to ensure samples were representative of the entire site area

Example # 2 – Effluent Guideline Study

- To support effluent guideline development, EPA samples raw source water, various in-process wastewaters, and final effluents from an industrial facility
  - Analyzes the samples a variety of pollutants using methods approved for CWA compliance monitoring
  - Methods encompass pollutants of interest to EPA as well as others
  - Lab contracts specify QC elements and QC acceptance criteria that must be met
  - Results are verified and validated. Anomalies are investigated and corrected (where possible) and impacts on the data quality are documented.
- Validated data provided to the data users

Example # 2 – Effluent Guideline Study (continued)

- **Study data users (EPA engineers and statisticians) plan to use data in evaluating waste treatment technologies**
  - Engineers notice presence of 2,4-diphenyl bad stuff in one of the in-process samples collected on Day 2
    - Not expected, based on known industrial processes or on published literature evaluated by EPA prior to study
    - Not present in same location on Day 1 or 3
    - Engineers review data validation reports and find no indication of a problem with the result. Ask the data reviewers to investigate further
    - Further review (including detailed review of raw data) yields no conclusive explanation for unanticipated presence of analyte
- **Data Quality Assessment Decision:**
  - Though nothing suggest the result is invalid, result is too questionable to rely on for rulemaking. Additional study is needed

Data Quality Assessment Tools

- EPA Science Policy Council *Assessment Factors* Guidance
- *Data Quality Assessment: A Reviewers Guide*, EPA QA/G-9R, February 2006
- *Data Quality Assessment: Statistical Methods for Practitioners*, EPA QA/G-9S, February 2006

General Assessment Factors

- **General assessment factors for evaluating the quality of scientific and technical information**
  - Soundness – The extent to which the procedures, methods, measures or models are consistent with the intended application
  - Applicability and utility – The extent to which the intended information is relevant for the Agency's intended use
  - Clarity and completeness – The degree to which the information can be understood and is sufficient for the intended use
  - Uncertainty and variability – Limitations on the reliability of results
  - Evaluation and review – The extent of independent verification, validation, and peer review of the information or of the procedures, measures, methods, or models

http://www.epa.gov/OSA/spc/pdfs/assess2.pdf

DQA – Five Steps

1. Review objectives and data collection design
2. Review verification and validation data
3. Select the statistical method for data evaluation
4. Verify the assumptions of the statistical method
5. Draw conclusions from the data

http://www.epa.gov/quality/dqa.html

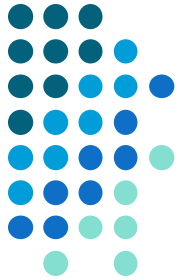
Data Quality Assessment Goes Beyond Standard "Data Review" – CLP Example

- CLP automated data review outputs are strictly binary functions
  - Does result fall within limits? Yes/No?
  - Do not provide interpretation of why a failure has occurred and what effect it may have on the specific result and overall project
  - Computer programs allow no grey area for further explanation or BPJ
- CLP NFGs require data validators to interpret the automated CLP data assessment output
  - "Bridge the gap" between the automated output and the assessment of data quality

Going Beyond Standard "Data Review" – Interpreting CLP flags

- Even with manual interpretation according to NFGs, additional data quality assessment is needed
  - Individual SDGs are evaluated as stand-alone products. (Does not evaluate all SDGs within a single project, or even site.)
- Example 1: Dual GC column analysis (e.g., pesticides, PCBs)
  - Both EXES and EPA ESAT chemists make no distinction between quality failures on one or both columns
  - Co-elution issues may be present on one column but not the other. A data quality assessor can determine if the automatically flagged data can actually be used.

Going Beyond Standard "Data Review" – Interpreting CLP flags



- Example 2: Matrix spikes
  - Automated CLP routines apply flags to matrix spike results that exceed NFG criteria.
  - Routines do *not* apply flags to associated sample results because NFGs require BPJ to evaluate impacts.
  - ☝ Note 1: If you are relying on the CLP flags without interpretation, you may be overlooking potential problems with your samples.
  - ☝ Note 2: This applies to many programs (not just CLP)

Data Suitability

- Suitability is a determination that the data review, verification, validation, and assessment have yielded the information necessary for the environmental decision
  - Suitability determination references all pertinent information from QAPP through assessment
  - Suitability determination arrives at the final conclusion that the data are, or are not, suitable for the environmental decision
    - If suitable, the decision is made
    - If unsuitable, it identifies changes or additional information necessary to inform the decision
  - Suitability determination is made by the Project Manager in concert with the QA Manager

Assessment Quiz

- True or false?
  - A statistical significance test may be necessary to determine if a compliance evaluation threshold has been exceeded?
  - Robustness is an assessment factor?
  - Validation data is all that is needed for making an environmental decision?
  - Data quality assessment is a 12-step recovery process?
  - All data flags have the same meaning?
  - All CLP data flags have the same meaning?
  - Suitability is a means of determining that Lou Blume is deserving of formal wear?

Agenda- Part 4

Part 4: Error Correction

Aw, c’mon...

- You’ve hung this long.
- We’re almost finished.

➢ We promise!!!

Estimated Time for Part 4:
5 minutes

Oops! We did everything right, but...

- Someone finds an error in an EPA database, at an EPA website, or in an EPA report or other document, the error can be reported:
  - Through EPA's Integrated Error Correction Process (IECP) (next slides)

http://www.epa.gov/enviro/html/frs_demo/dsman5_30_00.PDF

  - Through an Information Quality Guidelines (IQG) Request for Correction/ Request for Re-consideration RFC/RFR (later slide)

http://www.epa.gov/quality/informationguidelines/iqg-faqs.html#dataerror

EPA IECP

- Process by which the public can notify EPA of data errors
- Operates through “Envirofacts” web interface
- Provides a uniform mechanism and procedures for accepting input, routing, and tracking discrepancies
- Not intended to replace normal State or EPA error correction procedures
- Information gathered
  - Facility identification
  - Facility Registry System (FRS) identifying number, name, and location
  - Problem with the record as described by user
  - Solution proposed by user
  - User contact information

EPA IECP (cont.)

- Information routing
  - Data steward routes the issue to the appropriate program, State, or Regional office
  - Data steward remains the point of contact
- Communication
  - Data steward assures that communications are maintained until resolution
    - Communicates to user the plan to resolve the issue
    - Provides reasoning for resolution
    - Communicates resolution to user
- Tracking and reporting
  - Issues tracked in EnviroFacts
  - Monthly, quarterly, and yearly reports to OEI management

http://www.epa.gov/quality/informationguidelines/documents/EPA_InfoQualityGuidelines.pdf

http://oaspub.epa.gov/enviro/ets_grab_error.smart_form

IQG RFC/RFR Processes

- Request for Reconsideration (RFR) - if you don't like EPA's response to the Request for Correction (RFC)
- RFC/RFR processes provide a mechanism for the public to:
  - Request correction of information disseminated by a Federal Agency that does not comply with Information Quality Guidelines issued by OMB or the Agency
  - Challenge EPA information outside of the Administrative Procedure Act
- Recommend use of IECP for reporting a data error
- Additional information available via
  - Separate OW and GLNPO training modules
  - EPA IQG website

http://www.epa.gov/quality/informationguidelines/

Final Quiz (whew!)

True or false?

- The IECP operates through the Envirofacts web site?
- An IECP user must identify the facility at which he or she works?
- The “Data Steward” forwards the error to the appropriate State or Regional office?
- The RFC/RFR processes formalize error reporting?

Course Conclusions

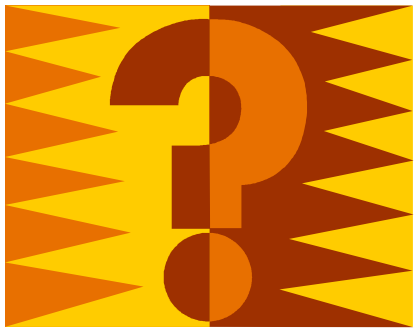
High quality data depends on up-front planning of:

- Project objectives
- Type of data needed for the project
- How the data will be used to support project objectives
- Quantity of data needed and how this need was determined
- Criteria for determining data quality and how those criteria were developed
- How when, and from where the data will be obtained, including existing (secondary) data
- How the data used in the project will be analyzed, evaluated and assessed

Course Conclusions (cont.)

- Data review, verification, validation, and quality assessment address the need to determine if data are suitable for making an environmental decision
  - Required by the Data Quality Act (aka Information Quality Act) and EPA's Quality System
  - If the data support the decision, all is well
  - If not, get back to work

Questions and Discussion



- In class, please use microphone
- On-line, please type your question

Feel free to ask a question or make a point based on your experience with data gathering, review, verification, validation, and data quality assessment

CONGRATULATIONS!

You have successfully completed the

Data Validation, Verification, and Usability Quality Management Training Module

If you would like to know more about QA or data review or have any questions, please contact:

- Your EPA QA Manager, QA Officer, or QA Coordinator http://www.epa.gov/quality/contacts.html
- EPA Quality Staff: quality@epa.gov or 202-564-6830
- Visit the EPA QA website at http://www.epa.gov/quality
- Marion Kelly, OW OST QA Officer, 202-566-1045, kelly.marion@epa.gov

EPA United States Environmental Protection Agency

Certificate of Completion

Webinar participants should contact Elizabeth Benjamin at ebenjamin@csc.com to receive a Certificate of Completion, and provide the following information:

- First and last name
- Affiliation *(if EPA, include Program Office)*
- Code word

Evaluation and Feedback

Please take a few moments to provide your feedback on this course.

Evaluation is online via
the Web-based program Survey Monkey at:

http://www.surveymonkey.com/s/DataVerification

Thanks for your participation!